

Essays in International and Urban Economics

Antonio Miscio

Submitted in partial fulfillment of the
requirements for the degree
of Doctor of Philosophy
in the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2016

©2016

Antonio Miscio

All Rights Reserved

ABSTRACT

Essays in International and Urban Economics

Antonio Miscio

Chapter 1, “The Impact of Trade Shocks on Local Labor Markets” estimates the effects of increased trade with China on Brazilian local labor markets using longitudinal individual data on the universe of Brazilian formal sector workers. First, I use reduced-form estimation strategies commonly found in the literature to compare my results to previous findings. I show that my results at the regional level mirror those found in prior studies based on cross-sectional data. I argue that these estimates are potentially biased as they do not take into account the flows of factors and goods between regions. I complement the reduced-form approach with a structural analysis based on the model by [Caliendo *et al.* \(2015\)](#) in order to endogenize such flows and to study welfare effects. I find that in the absence of the Chinese shock the Brazilian Commodities sector would have shrunk while Manufacturing and Services would have expanded. Relative to this baseline, the employment effect of increased trade with China at the national level was a slower reduction in the share of the Commodities sector and a slower growth in the Manufacturing subsectors that were relatively more exposed to Chinese import competition. My analysis suggests that while the average Brazilian worker benefitted from this shock, the welfare effects were very heterogeneous across sectors and across locations. I find that this heterogeneity is vastly underestimated if instead of using data at the level of metropolitan areas I use data aggregated by States and I explain why the choice of spatial units affects these results.

Chapter 2, “Agglomeration: A Long-Run Panel Data Approach” studies the sources of agglomeration economies in cities. We begin by incorporating within and cross-industry spillovers into a dynamic spatial equilibrium model in order to obtain a panel data estimating equation. This gives us a framework for measuring a rich set of agglomeration forces while controlling for a variety of potentially confounding effects. We apply this es-

timisation strategy to detailed new data describing the industry composition of 31 English cities from 1851-1911. Our results show that industries grew more rapidly in cities where they had more local suppliers or other occupationally-similar industries. We find no evidence of dynamic within-industry effects, i.e., industries generally did not grow more rapidly in cities in which they were already large. Once we control for these agglomeration forces, we find evidence of strong dynamic congestion forces related to city size. We also show how to construct estimates of the combined strength of the many agglomeration forces in our model. These results suggest a lower bound estimate of the strength of agglomeration forces equivalent to a city-size divergence rate of 1.6-2.3% per decade.

Chapter 3, “Gravity estimation with unobserved bilateral flow data” adapts the methodology by [Miscio & Soares \(2016\)](#) to predict domestic trade flows by sector between Brazilian metropolitan areas. This methodology, initially developed to infer *commuting* flows from aggregate data on population by place of residence and by place of work, relies on moment conditions derived from a general gravity equation and it is consistent with a large class of trade models. I show that it can also be applied to infer domestic trade flows by sector. Before using the methodology on Brazilian data, where we only observe flows between States, I test it on US data from the Commodity Flow Survey, where we observe both flows between States and between finer spatial units similar to metropolitan areas. I argue that the predicted bilateral flows obtained from this methodology are highly correlated with actual flows. Alternative approaches found in the recent literature differ from the one presented here in that they require stronger assumptions and deliver weaker results. In particular, the other approaches only describe aggregate flows (i.e. summing across all sectors) and cannot be used to predict sectoral flows.

Contents

List of Tables	iv
Lit of Figures	vi
Acknowledgements	vii
1 The Impact of Trade Shocks on Local Labor Markets	1
1.1 Introduction	2
1.2 Reduced-form analysis	6
1.2.1 Data and spatial units	7
1.2.2 Local exposure to increased trade with China	8
1.2.3 Regional evidence on employment, nominal earnings and migration	11
1.2.4 Individual-level evidence on employment and earnings	15
1.3 Theory	17
1.4 Structural analysis	20
1.4.1 Additional data sources and spatial units	20
1.4.2 From model to data	22
1.4.3 Main results	24
1.5 Conclusion	29
2 Agglomeration: A Long-Run Panel Data Approach	31
2.1 Introduction	32

2.2	Theory	37
2.2.1	Consumption	38
2.2.2	Production	39
2.2.3	Land and natural endowments	40
2.2.4	Timing	41
2.2.5	Production: Stage 3	42
2.2.6	Producers: Stage 2	43
2.2.7	Producers: Stage 1	45
2.2.8	Spillovers and technology diffusion	46
2.3	Data	49
2.4	Empirical approach	53
2.5	Main results	58
2.6	Strength of the agglomeration forces	64
2.7	Conclusion	71
3	Gravity estimation with unobserved bilateral flow data	73
3.1	Introduction	74
3.2	Methodology	75
3.3	Applications	81
3.3.1	Estimation of US sectoral domestic trade flows	81
3.3.2	Estimation of Brazilian sectoral domestic trade flows	86
3.4	Conclusion	91
	Bibliography	92
	Appendix	101

List of Tables

2.1	OLS and IV regressions including only one spillover path at a time	59
2.2	Results with all cross-industry spillover channels	60
2.3	Measuring the aggregate strength of the agglomeration forces	69
2.4	Measuring the aggregate strength of the convergence force associated with the within-industry effects	70
3.1	Cities in the primary analysis database	109
3.2	Industries in the primary analysis database with 1851 employment	111
3.3	Industry agglomeration patterns based on the Ellison & Glaeser index	112
3.4	Industry agglomeration patterns excluding London	113
3.5	Summary statistics for the cross-industry spillover terms	114
3.6	Correlations between cross-industry terms used in the main analysis	115
3.7	Simulated results with all parameters are set to zero vs. parameter esti- mates from true data	121
3.8	Weighted regression results with all cross-industry spillover channels	125
3.9	Features of industries that benefit from each type of cross-industry spillover	128
3.10	Features of industries that produce each type of cross-industry spillover	129
3.11	Features of industries that benefit from within-industry spillovers	130
3.12	Regression results with alternative functional forms	131

3.13	Alternative matrix regressions with one channel at a time	132
3.14	Regression results with cross-city variables	135

List of Figures

1.1	Surging Brazilian trade with China	3
1.2	Subnational variation in exposure to trade with China	4
1.3	Summary of reduced-form regressions at the regional level	12
1.4	An illustration of potential sources of bias	14
1.5	Incidence of localized shock on initial residents	16
1.6	Actual and predicted change in Brazilian imports from China per worker .	24
1.7	Predicted growth in Brazilian imports from China and growth in Chinese productivity implied by the model	25
1.8	Employment shares predicted by the model with and without China shock	27
1.9	Lifetime welfare effects by initial sector	28
2.1	Model timing	42
2.2	Strength of within-industry effects by industry	62
2.3	City size and city growth	67
3.1	Distance elasticities at the State and CFS-area level	83
3.2	Correlation between actual and predicted flows at CFS-area level	84
3.3	Correlation between actual and predicted flows at State level (Brazil)	88
3.4	Actual and predicted expenditure shares in Monte <i>et al.</i> (2015), sum across sectors	89
3.5	Actual and predicted flows, sum across sectors	90

3.6	Growth in real imports from China to Brazil and to other South American countries	102
3.7	Employment shares predicted by the model with and without China shock	103
3.8	Map showing the location of cities in the analysis database	110
3.9	Estimates and C.I.s from simulated results when all spillover parameters are zero	119
3.10	Simulated results with all parameters are set to zero vs. <i>IOin</i> estimate on real data	120
3.11	Robustness to dropping one city at a time – distribution of t-statistics	123
3.12	Robustness to dropping one industry at a time – distribution of t-statistics .	124

Acknowledgements

I thank my advisor Donald Davis for encouraging me since my first year of PhD and for guidance, assistance and advice; to my committee members Brendan O’Flaherty, David Weinstein, Jonathan Dingel and Walker Hanlon for discussions and feedback that influenced the content of this work; to seminar and conference participants at Columbia, at the annual meetings of the Urban Economics Association in Atlanta, of the European Urban Economics Association in Lisbon and the Spatial Economics Research Group at LSE; to my classmates and friends Ildiko Magyari, Evan Plous, Ilton Soares and Savitar Sundaresan for all their help. Above all, I thank my entire family for sharing this long ride with a smile on their faces.

All errors and omissions are my own.

Chapter 1

The Impact of Trade Shocks on Local Labor Markets

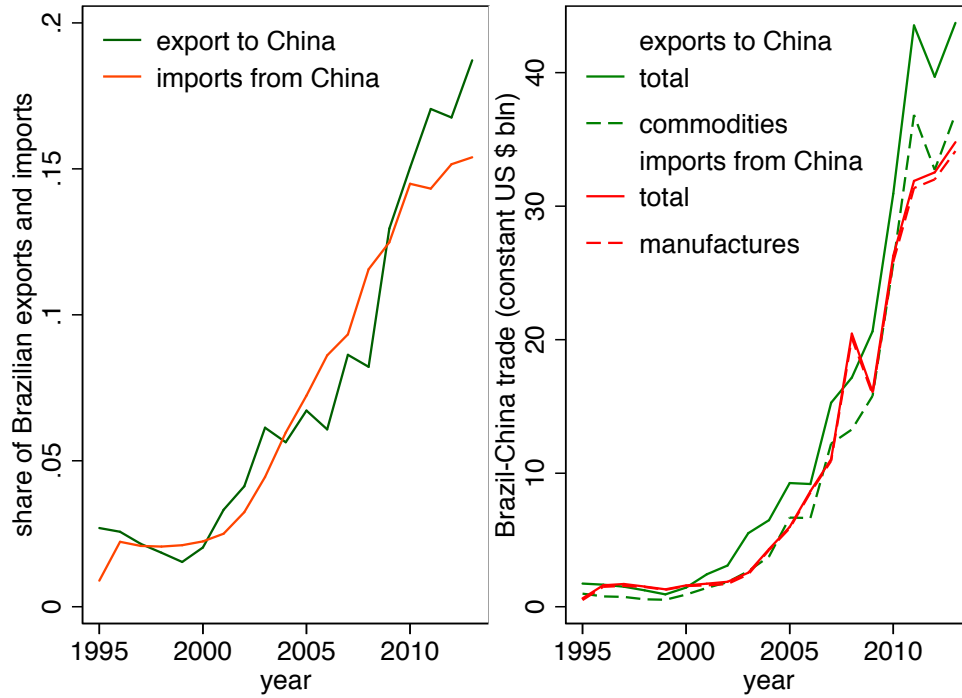
1.1 INTRODUCTION

What is the impact of a common policy shock on employment, *real* wages and migration across heterogeneous locations and across time? The three following observations make this a thought-provoking and challenging question. First, economic activity is heterogeneous across locations within a country. A consequence of this is that common policy shocks such as a trade liberalization have heterogeneous implications across locations. Second, spatial frictions matter. Therefore, shocks to one local market affect immediate neighbors more strongly than distant ones. Finally, the existence of non-spatial frictions implies that shocks take time to realize their full effect. While in isolation each of these facts seems intuitive, together they form an interesting and complex environment to study how local labor markets adjust to shocks. The answer to this question has important implications for policy design and evaluation. For instance, a better understanding of the implications of the Trans-Pacific Partnership at the subnational level would be useful to design effective Trade Adjustment Assistance programs.

The empirical context in which I answer this first-order question is in the regions of Brazil in the years following China's entry in the World Trade Organization (WTO) in late 2001. Unlike in the US, where trade with China was already on the rise in the 1990s, Brazilian trade with China was fairly flat and small in the years prior to China's WTO accession. As illustrated in figure 1.1, trade between these two countries was worth \$3 billion in 2000, representing just over 2% of Brazilian foreign trade. However, by 2013 these flows increased by a factor of twenty-six times in real terms. China is now Brazil's largest trading partner and it accounts for 17% of its trade. The composition of this bilateral trade reflects the two countries' comparative advantage: nearly all of Brazil's trade with China is in the exchange of commodities for manufactures.

At the national level and throughout the decade following China's WTO accession, bilateral trade between Brazil and China was fairly balanced. However, at the subnational

Figure 1.1: Surging Brazilian trade with China

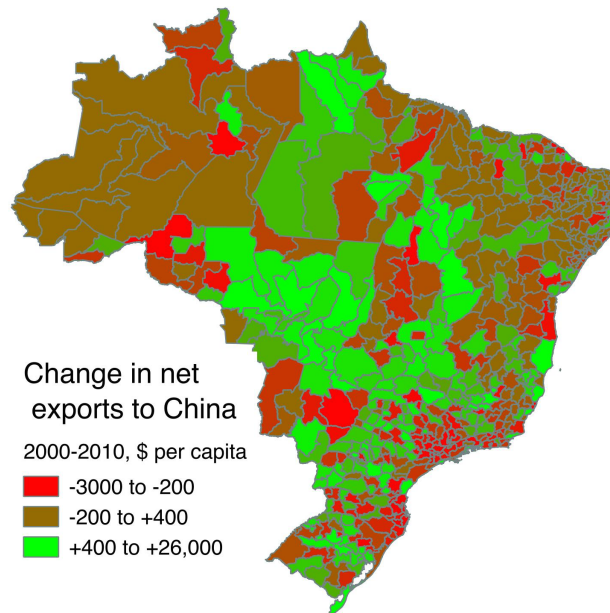


Notes: Trade between Brazil and China. In the left panel trade is measured as a fraction of total Brazilian exports and imports. In the right panel, it is measured in levels in constant 2010 US dollars (billions). Data: BACI.

level the picture looks very different. In figure 1.2 I use imports and exports data at the level of Brazilian municipalities to compute the local per-capita change in net exports to China. The map highlights that the implications of this policy change may have been very heterogeneous across locations due to pre-existing differences in regional industrial specialization. There is a sharp contrast between the commodity-rich areas of Brazil which saw a surge in exports, and the manufacturing regions which instead experienced serious competition from the flood of Chinese manufactured imports. These regional contrasts provide an ideal setting to answer my research question above.

In this paper, I first document the consequences of this large increase in trade with China on employment, *nominal* wages and migration across Brazilian regions and across time. The longitudinal nature of my data and its coverage of the universe of formal workers allow me to follow individuals across time and space and to describe labor market outcomes and migration histories with greater precision than what was possible in prior

Figure 1.2: Subnational variation in exposure to trade with China



Notes: The change in per-capita net exports to China between 2000 and 2010 (in constant 2010 US dollars) is computed using data on exports and imports by municipality aggregated to Micro-regions by the author. Population headcounts are from the 2000 Census.

studies that use cross-sectional data. While my results are consistent with intuition and in agreement with previous studies, I argue that the estimated impact of rising Chinese trade on US and Brazilian local labor markets may be biased. In particular, this is due to the assumption, implicit in many estimation strategies, that domestic local labor markets are exposed to foreign trade but that at the same time they are not connected to one another through flows of factors and goods. That is, paradoxically domestic regions are assumed to respond to trade shocks occurring at the national border and beyond but not to shocks occurring within the border. One of the ways to mitigate these identification issues is to focus on cohorts of workers rather than regions. Accordingly, I complement the reduced-form analysis at the regional level with a set of individual-level regressions aimed at assessing the extent to which this localized shock spread across space and time via migration. My results suggest that at a 10 year horizon up to 30 percent of the shock was absorbed through this diffusion channel.

Second, I argue that while data on nominal wages, employment and migration are

informative of local labor market adjustments (subject to the identification caveats discussed above), the assessment is at best incomplete and inconclusive without an analysis of *real* wages. Prior reduced-form studies have left this part of the question unanswered. Due to limited availability of data on local prices, in Brazil and elsewhere, a model is required to study real wages. An additional reason to introduce a model is the difficulty of taking into account flows of goods and factors in reduced-form regressions. Therefore, in the second part of this paper I borrow a dynamic model of trade from [Caliendo *et al.* \(2015\)](#) and apply it to Brazilian data. Some of the assumptions in this model make it more suitable to study integrated labor markets than larger regions such as States. Yet, data limitations prevented the authors from taking their model too seriously and they apply it to study the impact of increased trade with China on US States. Hence, this is the first paper to estimate the model in [Caliendo *et al.* \(2015\)](#) using data at the level of metropolitan areas. I find that, overall, the welfare of Brazilian workers improved as a result of increased trade with China. However, I also find substantial variation across sectors and regions. In addition, I re-estimate the model using State-level data and discuss the implications of using data at a higher level of spatial aggregation. My results show that from a national perspective it does not matter whether I use data at the level of States or metropolitan areas. However, I show that by using spatially aggregate data I vastly underestimate the amount of heterogeneity occurring at the subnational level. This suggests that there is substantial heterogeneity within States and that this richness of detail is simply lost (or averaged out) in estimations at more aggregate levels.

This project is related to several literatures. A strand of papers has investigated the effects of trade shocks on local labor markets through comparative statics and using decade-long differences. This includes [Topalova \(2007\)](#), [Autor *et al.* \(2013\)](#), [Kovak \(2013\)](#) and [Costa *et al.* \(Forthcoming\)](#), among others. Another branch of literature has looked at the impact of other types of localized shocks through the lens of a spatial equilibrium model. In particular, [Diamond \(Forthcoming\)](#), [Notowidigdo \(2013\)](#) and [Yagan \(2014\)](#) belong to

this group. The closest paper in terms of reduced-form regional analysis and data sources is [Dix-Carneiro & Kovak \(2015\)](#), who study the impact of Brazil’s trade liberalization in the 1990s at yearly frequency. Like most of the papers referenced above, they also omit to analyze the impact on local prices and real wages.

Finally, this paper builds on a more structural literature that estimates dynamic models of workers’ migration choice ([Kennan & Walker \(2011\)](#), [Bishop \(2012\)](#) and [Schmutz & Sidibe \(2015\)](#) among others) and the structural literature that studies the impact of trade shocks on labor market dynamics ([Artuc *et al.* \(2010\)](#), [Dix-Carneiro \(2014\)](#), [Caliendo *et al.* \(2015\)](#) and references therein).

In section [1.2](#) I present the main data sources, I discuss the spatial variation in exposure to trade with China across Brazilian regions and present a reduced-form empirical analysis at the regional and individual level. In section [1.3](#) I present a summary of the theory by [Caliendo *et al.* \(2015\)](#). In section [1.4](#) I present additional data sources used in the structural analysis and the main results from this estimation. Finally, section [1.5](#) concludes.

1.2 REDUCED-FORM ANALYSIS

I begin by presenting the data sources used in the reduced-form analysis and my choice of spatial units. Then, I introduce a commonly used measure of regional variation in exposure to trade and explain how to interpret the rise in trade with China as a plausibly exogenous shock to Brazilian local labor markets. I follow a large reduced-form literature on local labor markets by using a differences-in-differences approach (DD) to document the response of employment, nominal earnings and migration at a higher frequency than in prior studies, which allows me to study not only the effects at a decadal horizon, but also how they change year by year. I discuss the sources of bias that could potentially affect these estimates as well as those found in prior studies. Finally, I turn my attention to individual-level regressions in order to avoid some of the identification concerns in-

herent in the regional-level analysis and to assess the extent to which this localized shock diffused across space and time through migration.

1.2.1 Data and spatial units

The choice of an appropriate spatial unit is fundamental in a study that focuses on *local* labor market effects. An influential study on the effects of increased trade with China on US local labor markets by [Autor *et al.* \(2013\)](#) uses the Commuting Zones defined by [Sizer & Tolbert \(1996\)](#).¹ A follow-up study on Brazilian data by [Costa *et al.* \(Forthcoming\)](#) uses Micro-regions, which are defined by the Brazilian Institute for Statistics and Geography (IBGE) as an aggregation of Municipios that have similar productive structures. Several other studies on Brazilian local labor markets, including [Dix-Carneiro & Kovak \(2015\)](#), also use Micro-regions. They justify their choice by arguing that the average commuting rate between Micro-regions is less than 5 percent. However, as I find in a separate study available on request, this average hides the fact that some Municipios send a much larger share of their workforce to locations outside of their Micro-region, and that the vast majority of Micro-regions consists of Municipios that are not connected by strong commuting ties. Rather than using off-the-shelf spatial units that are not designed to capture integrated labor markets, my solution is to construct commuting zones from the bottom-up by applying the methodology developed in [Duranton \(2013\)](#) to Brazilian commuting data.

Therefore, I use micro-data on workers' residence and workplace locations from the 2010 Census to compute commuting flows between Brazilian Municipios and then I use these flows to aggregate them into metropolitan areas. The resulting spatial units are integrated labor markets that by construction have strong commuting ties within and weak

¹[Monte *et al.* \(2015\)](#) find that there are significant commuting flows across these Commuting Zones. However, the assignment of Counties to Commuting Zones in [Sizer & Tolbert \(1996\)](#) is based on 1990 commuting data, while [Monte *et al.* \(2015\)](#)'s assessment is based on data from 20 year later. A fair assessment requires the use of data from the same period. See next footnote.

ones across.² In what follows, I use *location*, *region* and *metropolitan area* as synonyms for an integrated local labor market defined as I just described. In the reduced-form analysis I focus on the 200 largest metropolitan areas, which roughly corresponds to a minimum population cutoff of 100,000 people. These metropolitan areas represent over 60 percent of Brazil's total population in 2010.

My main data source in the reduced-form analysis is the Brazilian *Relação Anual de Informações Sociais*, a matched employer-employee dataset provided by the Ministry of Labor and Employment, for the years 2002 to 2013. This confidential dataset contains job records of the entire Brazilian formal sector. In particular, it contains worker and establishment identifiers, which allow me to follow workers as they move across employers, across space and to follow them over time. The main variables of interest for this study include workers' demographics, occupation, education and nominal earnings, as well as their employers' industry and municipality.

Additional micro-data comes from the Brazilian Population Census for the years 1991, 2000 and 2010. This is mainly used to measure regional labor market outcomes prior to China's entry in the WTO.

1.2.2 Local exposure to increased trade with China

A common approach to gauge regional variation in exposure to international trade is to interact a national trade shock with regional industry shares, as in [Topalova \(2007\)](#). I follow [Autor et al. \(2013\)](#) in defining local exposure to Chinese trade as a locally weighted average of changes in imports and exports per worker. More formally, I measure trade exposure as $\Delta M_c = \sum_i \frac{\Delta M_i}{L_i} \cdot \frac{L_{ic}}{L_c}$ and $\Delta X_c = \sum_i \frac{\Delta X_i}{L_i} \cdot \frac{L_{ic}}{L_c}$, where ΔM_i is the change in

²In a separate study, available on request, I use US commuting data from 2010 to update [Sizer & Tolbert \(1996\)](#)'s definition of Commuting Zones and to construct an alternative mapping from Counties to commuting-based metropolitan areas à la [Duranton \(2013\)](#). I find that both mappings result in a distribution of log population that is highly correlated with that of Core-based Statistical Areas (CBSA) defined by the US Office of Budget and Management. The correlation ranges between 0.90 and 0.98 depending on the commuting thresholds. This exercise provides some assurance that, although Brazil lacks the equivalent of the US CBSA, an alternative definition based on commuting flows is a good substitute.

Brazilian imports from China in industry i between 2000 and 2005, ΔX_i denotes changes in Brazilian exports to China, L is employment and c indexes regions. The first factor captures changes in industry exposure to foreign trade at the national level, while the second factor measures local variation by weighing industries according to their share of local employment. Therefore, the cross-sectional variation in trade exposure is entirely driven by differences in the regional industry mix in 2000.³ In order to address concerns of endogeneity and reverse causality, in all the regressions presented here I instrument these changes in bilateral trade between Brazil and China with changes in trade between China and all South American countries other than Brazil. This is analogous to Autor *et al.* (2013)'s use of the change in imports from China by other developed countries as an instrument for the change in US imports from China.

Even though I observe foreign imports and exports by subnational units, which I use for illustrative purposes in figure 1.2, in the rest of the analysis I prefer to use a measure of local exposure to trade that depends on local employment shares. This is because, for example, a region may still be exposed to import competition even though it only sells domestically both before and after the shock. Data on imports and exports by subnational units would fail to capture this aspect. Nonetheless, in my sample period the two measures are highly correlated.

As shown in figure 1.2, there was significant geographical variation in the degree of exposure to trade with China.⁴ In particular, there are two points worth emphasising. First, the sign and the magnitude of the change in per-capita net exports to China in the period 2000-2010 varies across regions. Regional variation in trade exposure analogous to this one has been used in most of the reduced-form studies cited earlier. However,

³The choice of 2005 as the end period is inconsequential. Choosing 2010, or any other year in between simply leads to different point estimates, but it does not alter the significance of the estimated coefficients or their relative magnitudes. This is because in all the reduced-form regressions presented here the variation is cross-sectional, i.e. the employment shares are fixed at their 2000 value and the trade flows changed in magnitude over time but not in their sectoral composition.

⁴While in the rest of the analysis I use commuting-based metropolitan areas at the unit of observation, in this map it is also appropriate to use Micro-regions since these are defined as sets of Municipios that have similar production structures.

consider two areas with the same sectoral employment shares and therefore the same exposure to trade with China as define above, but such that one of them is surrounded by areas that experienced a positive change in net exports and such that the other one is surrounded by areas that experienced a negative change in net exports. Intuitively, these two regions are *indirectly* exposed to different trade shocks even though the *direct* shock is identical by construction. Shocks to nearby regions may in principle spill over and reinforce, or counterbalance the ones directly affecting any one region. This second feature of the data is another source of regional variation in the degree of exposure to Chinese trade, albeit one that has been ignored in most prior studies. This amounts to assuming, implicitly, that these regions are disconnected from one another, i.e. not linked by trade or migration flows.

In a contemporaneous study, [Monte *et al.* \(2015\)](#) make a similar point by arguing that local employment elasticities vary across locations as they depend on the degree of interaction between any location and its neighbouring regions as measured by commuting, trade or migration flows. Their estimation on US Counties suggests that most of the variation in local employment elasticities depends on differences in commuting links and they propose a model-based correction to account for such differences in reduced-form regressions. As discussed earlier, the spatial units in my reduced-form analysis are integrated labor markets constructed using a common minimum commuting threshold. This eliminates most of the variation in commuting links with neighbouring regions. Yet, migration and trade flows are not used in the construction of my spatial units so [Monte *et al.*](#)'s general point remains valid.

In the following section, I document the impact of increased trade with China on regional labor outcomes at different time horizons using the variation in regional trade exposure as measured by the indicators presented above. In the interest of comparing my results with previous studies, I follow a traditional DD approach (i.e. without the corrections proposed by [Monte *et al.* \(2015\)](#)) and then I show that some of the estimated

elasticities change once we take into account the shock occurring in nearby regions.

1.2.3 Regional evidence on employment, nominal earnings and migration

I use a DD approach similar to [Dix-Carneiro & Kovak \(2015\)](#) to compare regions that were on parallel growth trajectories prior to the shock (conditional on the controls) but experienced differential exposure to trade with China. The equation I estimate takes the form

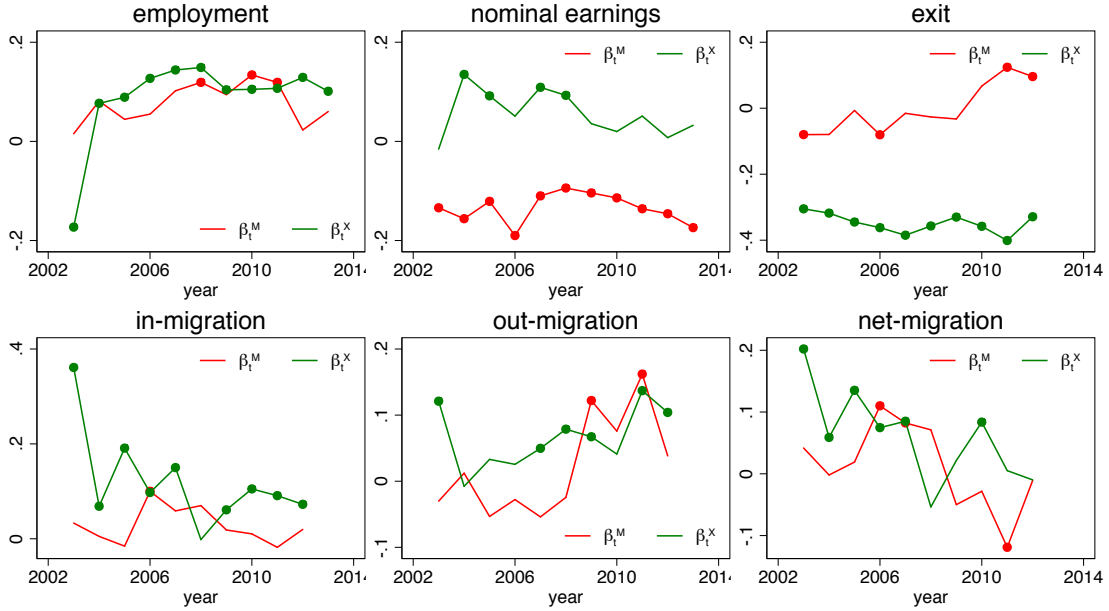
$$\Delta y_{c,t-2002} = \alpha_{st} + \beta_t^M \Delta M_c + \beta_t^X \Delta X_c + \delta_t \Delta y_{c,2000-1991} + \varepsilon_{ct} \quad (1.1)$$

where $\Delta y_{c,t-2002}$ is the change in the outcome of interest (log employment, log nominal earning premia, exits from the formal labor force, in-migration, out-migration and net-migration) in region c between year t and the base year 2002, α_{st} is a state-time fixed effect and $\Delta y_{c,2000-1991}$ is a pre-trend in the outcome of interest.⁵ Although this specification could lead to biased estimates, as I explain below, I use it in order to compare my results to the existing literature and to establish a benchmark.

Figure 1.3 shows a summary of all regressions at the regional level in graphical, rather than table format. Overall, the pattern of results is consistent with the intuition that workers faced better conditions in the regions specialized in export industries than in those competing with Chinese imports. For instance, I estimate significant differences in the change in nominal earnings between exporting and import-competing regions. Similarly, the results suggest that in-migration of formal workers was significantly higher in exporting regions and exits from the formal labor force significantly less prevalent there than in import-competing regions. These results also mirror those found by [Costa et al. \(Forthcoming\)](#) who use cross-sectional Brazilian census data. What is new here is the time

⁵Log nominal earning premia are the regional fixed effects computed from a Mincerian regression of workers' log nominal earnings on demographics, education, industry and region dummies. In regressions where the dependent variable is log nominal earning premia, observations are weighted to correct for heteroskedasticity.

Figure 1.3: Summary of reduced-form regressions at the regional level



Notes: Each point is estimated in a separate regression as in equation 1.1 . Circles denote that the coefficient is significantly different from zero. All variables are standardized, so the coefficients represent the differential effect of one standard deviation more exposure to trade with China than the average region.

profile of these effects in the inter-census years. I find, for instance, that in-migration and net-migration to exporting regions is much higher at the beginning of the decade than in subsequent years.

While these estimates are in agreement with previous studies and consistent with intuition, their interpretation as the effect of increased trade with China is open to debate. Such interpretation is not valid if workers reallocate across regions in response to the shock. The same criticism potentially applies to [Topalova \(2007\)](#), [Autor *et al.* \(2013\)](#), [Kovak \(2013\)](#) and [Costa *et al.* \(Forthcoming\)](#) and all those who implicitly treat local labor markets as disconnected islands. [Baum-Snow & Ferreira \(2015\)](#) discuss in detail the threats to identification when using DD in settings where the data is aggregated - e.g. regions - and individuals can re-sort between treatment and control groups. Failure to control for this spatial spillover results in an omitted variable bias. A similar, but more general point is made by [Monte *et al.* \(2015\)](#), who argue that interpreting these results as the coefficient on the treatment is not warranted if the control group is also affected

by the treatment through spatial linkages in factor and goods markets. It is more general than what argued by [Baum-Snow & Ferreira \(2015\)](#) because, in the hypothetical scenario where individuals are immobile, [Monte et al. \(2015\)](#)'s point is that shocks to one region can in principle spill over to other regions via trade flows. As shown in the bottom panel of figure 1.3, I find indeed evidence of endogenous sorting of workers in response to the same shock. This evidence makes me question the interpretation of the results in the top panel in the same figure as the effect of local exposure to increased trade with China.

More formally, suppose that the data generating process is the following⁶

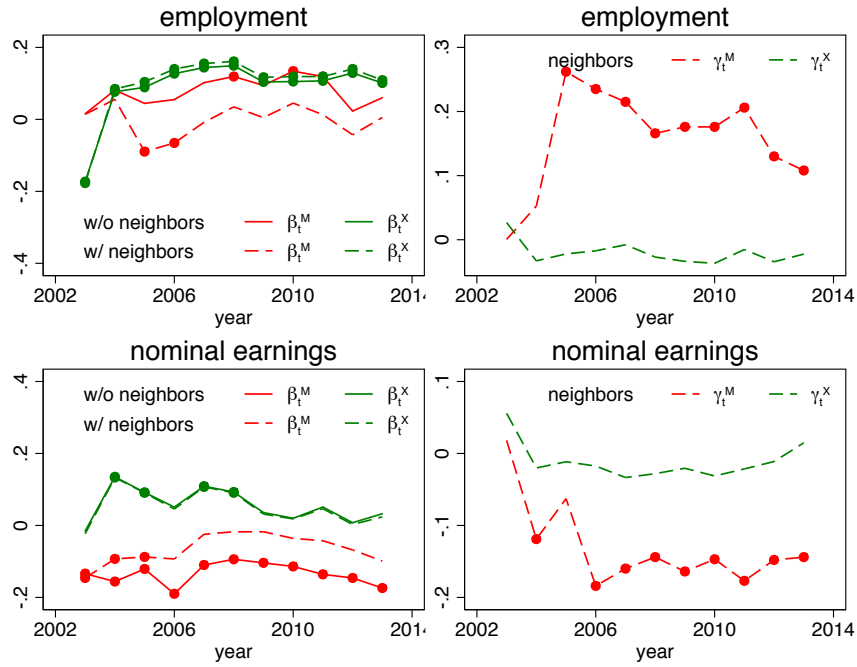
$$\Delta \log(\text{employment})_{c,t-2002} = \beta_t^M \Delta M_c + \gamma_t^M \Delta \tilde{M}_c + \varepsilon_{ct} \quad (1.2)$$

where β_t^M is the coefficient on the direct import shock in region c (denoted by ΔM_c), and γ_t^M is the coefficient on the indirect import shock (denoted by $\Delta \tilde{M}_c$ and defined as some function of the import shock occurring in nearby regions). The sign of the bias resulting from omitting $\Delta \tilde{M}_c$ depends on the product between γ_t^M and the covariance between ΔM_c and $\Delta \tilde{M}_c$. It is easy to think of scenarios where a shock in neighboring regions affects labor demand in region c via lower demand for c 's products (hence $\gamma_t^M < 0$), or it may affect labor supply in region c via higher net-migration than in the absence of a shock in the neighboring regions (hence $\gamma_t^M > 0$). In summary, the estimated coefficient of interest may be higher or lower than the actual β_t^M since the sign of the bias depends on which effect dominates.

This also means that the estimated effects shown in figure 1.3 could be different from the true coefficients. As an illustration, I re-estimated the equations where the dependent variable is log employment and log nominal earnings with two additional controls analogous to $\Delta \tilde{M}_c$ from equation 1.2. In particular, I computed a distance-weighted average of the import and export shock in the five closest regions. I report the results in figure 1.4. The solid lines in the left panel are identical to the solid lines in figure 1.3. The dashed

⁶I omit the additional terms from equation 1.1 to simplify the exposition.

Figure 1.4: An illustration of potential sources of bias



Notes: In the left panel I estimate the coefficient on a region's own import shock with and without the inclusion of the shock occurring in neighbouring region as additional controls. In the right panel I estimate the coefficient on the additional controls. The shock occurring in nearby regions is measured as a distance-weighted average of the import and export shock in the five closest neighbouring regions.

lines in the left panel are the coefficients on ΔM_c and ΔX_c after we introduce the additional controls. The estimated effect of a region's own import shock on nominal earnings is now not distinguishable from zero at a ten-year horizon and we do find a significant effect of an import shock in the nearby regions. An import shock in the nearby regions appears to have a positive effect on employment and a negative effect on earnings, both of which are consistent with an increase in the local labor supply via in-migration.

The conclusion so far is that without a model that accounts for general equilibrium interactions between regions, regional effects of trade shocks are hard to estimate. In addition, with cross-sectional data the identification of the coefficients of interest is even harder since individuals can potentially sort between treatment and control groups in response to the same shocks. In the next section I turn to individual-level regressions to mitigate some of these identification concerns inherent in regional-level regressions.

1.2.4 Individual-level evidence on employment and earnings

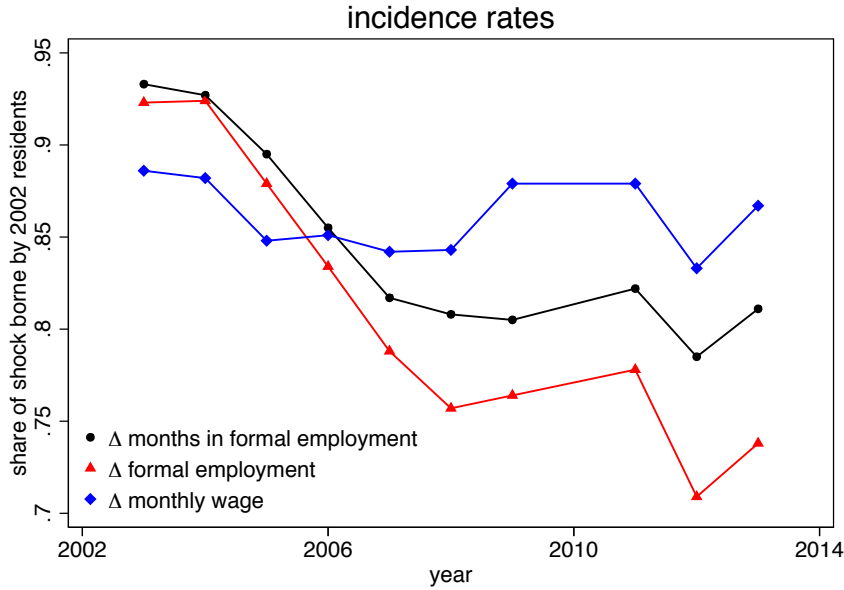
Since my main dataset is longitudinal at the individual level, I could choose to follow specific cohorts of workers and attempt to estimate the impact of increased trade with China on those initially living in a given location. This is analogous to the identification strategy in [Autor *et al.* \(2014\)](#). I would solve the problem faced by researchers who use cross-sectional data and cannot distinguish whether an observed labor outcome is due to the shock or to a change in the composition of the local labor force. However, this reduced-form analysis at the individual level would still not solve the issue raised by [Monte *et al.* \(2015\)](#). That is, I would still not be taking into account that shocks can spill over between treatment and control groups via trade flows. Let us consider a specific example to illustrate this point. Suppose that (a) import competition in a given location is 10 percent higher than in the average region, and (b) that cumulative earnings for the cohort of workers initially living in that location is 10 percent lower than for the average cohort. We still cannot quantify the effect of local trade exposure on that cohort because in order to identify it we would also need to know what fraction of the reduction in cumulative earnings is due to shocks occurring elsewhere and spilling over via trade flows.

However, my ability to observe individual migration histories can still be useful in quantifying the incidence of local shocks on initial residents using the identification strategy proposed by [Yagan \(2014\)](#). His approach is based on a comparison of labor outcomes between (a) the cohort initially living in a location and (b) the workforce living in that location in subsequent years. More formally, I estimate the following equation:

$$\Delta y_{it} = X_i \beta_t + \gamma_t \Delta \bar{y}_{c(i,t_0)t} + \varepsilon_{it} \quad (1.3)$$

where y_{it} is worker i 's outcome of interest in year t (formal employment status, monthly earnings, months in formal sector), X_i is a vector of individual controls, $\bar{y}_{c(i,t_0)t}$ is the

Figure 1.5: Incidence of localized shock on initial residents



Notes: Coefficients from individual-level regressions on all workers who in 2002 were in the formal sector and aged 25-55.

average outcome in i 's initial location in period t (i.e. her residence location at time t_0) and Δ denotes the change between 2002 and $t \in [2003, 2013]$. By construction, if nobody migrates then labor outcomes measured on groups (a) and (b) are perfectly correlated. On the other hand, if workers migrate to locations with better economic prospects, then that correlation weakens.

I follow Yagan (2014) in interpreting the value of γ_t as the share of shock borne by 2002 residents. As shown in figure 1.5, initial residents bear 80-85 percent of the local shock after 5 years and 70-80 percent after 10 years. For comparison, in the US Yagan (2014) finds that the incidence of the Great Recession on initial residents was 93 percent after 5 years. He argues that local labor markets were depressed everywhere across the US (albeit to a different degree) and this provided weak incentives to migrate. On the other hand, in the Brazilian context, the contrast between exporting and import-competing regions seems to have provided much stronger incentives to migrate and to smooth local idiosyncratic shocks.

While the upper bound on γ is clear, it is not obvious what the lower bound should

be. In other words, a reduced-form exercise like this one suggests that migration is partly responsible for the diffusion of localized shocks across time and space, but it remains unclear how much further γ should drop until a local shock has been completely absorbed. This indeterminacy is related to the fact that *real* wage differentials drive individual migration decisions, not nominal ones. Therefore, it is possible that monthly wages do not need to drop any further if local prices absorbed the remaining differentials. However, without detailed price data we cannot assess the effect of increased trade with China on local price indexes. This is the second reason why a model would be useful to answer my initial question.

To summarize this section, I have estimated reduced-form effects of increased trade with China at the regional and individual level using estimation strategies drawn from the existing literature. I have argued that the existence of inter-regional flows of goods and factors makes the identification of such effects very problematic. The availability of longitudinal individual data partly mitigates this problem but it does not solve it entirely. In the next section I will introduce a model with the purpose of taking into account such inter-regional flows and of conducting a welfare analysis despite the lack of detailed price data.

1.3 THEORY

In this section I use a dynamic trade model proposed by [Caliendo *et al.* \(2015\)](#) to complement the reduced-form analysis presented above. As I argued earlier, a model is needed to address two shortcomings of the reduced-form approach. The first shortcoming I discussed is the difficulty of estimating reduced-form regional or individual effects that take into account the existence of inter-regional flows of factors and goods. Second, in the absence of detailed price data I cannot study the effect of increased trade with China on local prices in a reduced-form fashion, and therefore I cannot discuss welfare effects.

The model by [Caliendo et al. \(2015\)](#) solves both problems.

I provide here a summary of the model in order to build some intuition for the empirical application. I refer the reader to the original paper for a full derivation and for additional details. The economy consists of multiple regions and multiple sectors linked to one another through input-output connections and through trade and migration flows. Each region is inhabited by forward-looking households who consume final goods and supply labor to firms. Immobile rentiers also consume final goods and supply structures to firms. Production is done by a continuum of perfectly competitive firms that combine labor, structures and material inputs using a technology with heterogeneous productivity as in [Eaton & Kortum \(2002\)](#). Materials are simply final goods produced by other firms that get recombined with labor and structure to produce new final goods. While at the beginning of each period households take their location and sector of employment as given, they can choose to reallocate across locations and sectors at the end of the period after incurring a reallocation cost and according to an *i.i.d.* taste shock.

I have two main reasons for preferring this model over existing alternatives. First, it provides a very detailed and flexible description of costly inter-regional flows of goods and factors and it is dynamic in nature. Second, despite its rich features, this model remains tractable and amenable to empirical research thanks to [Caliendo et al.](#)'s innovative solution method based on time differences. In particular, they show that all the static equilibrium conditions can also be written in relative changes, which means that all time-invariant parameters cancel out and need not be calibrated or solved for. This solution concept drastically reduces the computational burden required to find a full solution and makes this a feasible empirical model.⁷ For instance, the market clearing condition for

⁷For comparison, in an application with J sectors, N countries and R regions in each country if the equilibrium conditions are not written in relative changes, a researcher would need to estimate $N \times R \times J$ productivity levels and as many stocks of local structures, $N^2 \times R^2 \times J^2$ asymmetric bilateral trade costs and as many labor mobility cost. On the other hand, by expressing the same equilibrium conditions in time differences the only parameters needed to solve the model are the ones that are not time-invariant. In the application presented in this paper, the only parameters assumed to change over time are the Chinese productivity levels.

structures (in levels) is the following:

$$H^{nj} = \frac{\gamma^{nj} \zeta^n}{r_t^{nj}} \sum_i^N \pi_t^{ij,nj} X_t^{ij} \quad (1.4)$$

where H^{nj} is the time-invariant stock of local structures in location n and sector j , γ and ζ are respectively the output share of value added and the share of structures, r is the rental rate of structures, X^{ij} is total expenditure in location i on goods from industry j and π is the trade share in i on goods from n , i.e. the fraction of X_t^{ij} spend on goods imported from location n . Even if all the parameters in this equation were time-invariant, we would need to know or estimate all of them in order to solve for the endogenous variables using this equilibrium condition. However, by writing it in time-differences as follows, all the time-invariant parameters cancel out:

$$\hat{r}_t^{nj} = \left(\sum_i^N \pi_t^{ij,nj} X_t^{ij} \right) / \left(\sum_i^N \pi_{t-1}^{ij,nj} X_{t-1}^{ij} \right) \quad (1.5)$$

where $\hat{r}_t = \frac{r_t}{r_{t-1}}$.

[Caliendo et al. \(2015\)](#)'s model is formulated in terms of abstract subnational *regions*, which in empirical applications could in principle correspond to any level of spatial aggregation. Indeed, in their own empirical application on US data [Caliendo et al.](#) choose US States. However, the modelling of local congestion forces that raise the cost of living suggests that a natural application of this model should be on metropolitan areas, rather than States. In the authors' own words (p. 66) "migration has a negative effect on real wages because the inflow of workers strains local fixed factors and raises the relative price of structures and the cost of living". The choice of a spatial unit is not an innocuous one because it implicitly affects the scale at which all mechanisms in the model operate. The following example illustrates why I believe that this model is a better description of metropolitan areas than States. Suppose that a State contains two metropolitan areas, one specialized in a declining industry and the other one specialized in a growing industry.

Let us assume that geographical and sectoral relocations are costless but only allowed within the State, that the price of traded goods is set internationally and that the cost of living depends only on congestion. Economic intuition suggests that over time a fraction of the workforce may migrate between the two metropolitan areas, and that the cost of living in the two places will move in opposite directions. However, if we analyse this model economy at the State level, by construction we do not see any migration and therefore the model would predict no change in the cost of living, i.e. it would under-estimate both the decline in the cost of living in one area and the increase in the other. In my application to Brazilian data, I estimate this model at both spatial scales and I show which results are affected by the level of spatial aggregation in the data.

1.4 STRUCTURAL ANALYSIS

1.4.1 Additional data sources and spatial units

Although the model in [Caliendo *et al.* \(2015\)](#) can accomodate an arbitrary number of regions, I follow their empirical application by limiting the number of foreign regions to one per foreign country. The number of distinct foreign countries included in the analysis is driven by the coverage of the World Input-Output Database, which contains data on 40 major countries and a combined Rest of the World (RoW). I merge 23 of these countries into the RoW as they individually represent less than 1 percent of Brazilian trade in 2000 and collectively less than 5 percent. This leaves me with 17 foreign regions in addition to Brazil. In Brazil, I perform a separate analysis for each of two levels of spatial aggregation. First, I divide the country in States. Due to data limitations, I only include in the analysis 26 of Brazil's 27 States. The omitted one, Acre, contains less than 1 percent of Brazil's total population. Then, I divide Brazil into the union of commuting-based metropolitan areas, as defined in section [1.2.1](#), and State residuals. In order to reduce the number of parameters to be estimated, I increase the minimum population cutoff relative to the

reduced-form analysis from 100,000 to 150,000. This leaves me with 100 metropolitan areas which jointly contain 55 percent of total population. I merge all remaining Municipios into 25 State residuals.⁸

The data required to estimate this model include bilateral trade flows between all regions (whether domestic or foreign), value added by industry and an input-output matrix for each region, the distribution of employment by industry-region pairs and migration flows between all sectors and regions. All of the above is only needed for the initial period and, as required by model assumptions, the data should come from a period that precedes the shock we are interested in studying.

The Brazilian RAIS dataset, already presented in section 1.2.1, is particularly useful in that it allows me to observe migration rates between Municipio-industry pairs. In the US application by [Caliendo *et al.* \(2015\)](#) the lack of such data is what prevented the authors from estimating their model at a more disaggregate level than States. In fact, even at the State level they had to make a number of assumptions to obtain the complete matrix of migration flows between industry-location pairs.⁹

As discussed in more details in chapter 3, trade data at the subnational level (i.e. trade flows between subnational units) is hard to obtain in most developed and developing countries. The US Commodity Flow Survey (CFS) is a rare exception. In Brazil, the only domestic trade data that I was able to find comes from [de Vasconcelos & de Oliveira \(2006\)](#) and it contains trade flows between most Brazilian States by industry for the year 1999. In chapter 3 I show how to use gravity equations and data on production and consumption by location to estimate domestic sectoral trade flows. This methodology allows me to both complete the matrix of trade flows at the State level, for States not included in

⁸The total number of States is 27. The State of Acre is dropped for data limitations. The State-equivalent territory of Brasilia is entirely contained in the Brasilia commuting-based metropolitan areas. Therefore, there are only 25 State residuals.

⁹For instance, [Caliendo *et al.* \(2015\)](#) compute transitions between industries using data from the Current Population Survey (CPS) and transitions between States using the American Community Survey (ACS). However, the CPS only follows individuals who reside at the same address so the authors have to assume that individuals who move across States have the same sectoral transition probabilities as those who do not change address at all.

de Vasconcelos & de Oliveira (2006), and to infer trade flows between metropolitan areas, on which there is no data. In the same chapter, I validate the methodology by applying it first to US CFS data. I show that by combining an industry-specific distance elasticity estimated at the State level with data on sectoral production and consumption by CFS regions (similar to metropolitan areas), I am able to compute predicted flows whose correlation with actual flows is around 0.80. I refer the reader to chapter 3 for more details about this methodology and for a discussion of potential alternatives (or lack thereof).

The rest of the data is from standard sources. I obtain bilateral trade flows between countries and national input output matrices from the the World Input Output Database (WIOD) for the year 2000, my base year. Since for Brazil I lack value added data at the subnational level, I use value added shares by industry at the national level from the WIOD and assume that they are identical across subnational units. Data on employment by industry at the local level is from the Brazilian census in 2000.

The sectoral classification I use in the analysis is based on the one in the WIOD. I collapse some of the sectors in order to reduce the parameter space, which is particularly useful in the structural analysis at the metropolitan area level.

1.4.2 From model to data

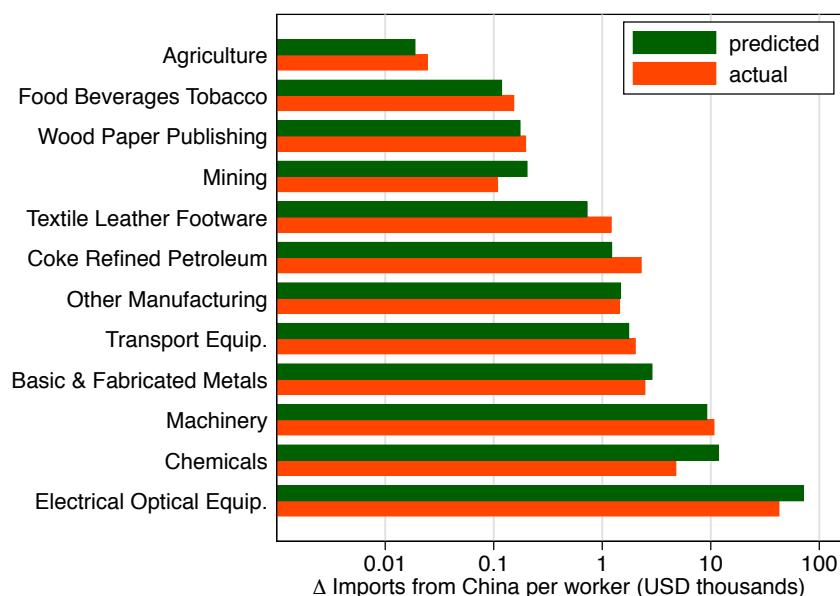
In this section I explain how I take the model presented in the previous section to the data. Caliendo *et al.* (2015) show how to solve their model numerically given data from the initial period. In particular, there are two versions of the model that we may want to solve. In the first version we assume that there is no shock. Therefore, given the equilibrium conditions written in relative changes, the data from the first period is sufficient to solve the model in all subsequent periods. In the second version, we assume that a specific shock occurs after the first period and that it affects the parameters in the model in a known way. Then, we can use the solution computed under the assumption that there are no shocks and the new equilibrium conditions that incorporate the shock to

solve for the new path of endogenous variables.

For the first version of the model, I assume that no shock occurred before the year 2000, my base year. For the second version, I calibrate the change in Chinese productivity between 2000 and 2010 using a strategy similar to [Caliendo *et al.* \(2015\)](#). First, I compute the growth in imports from China to Brazil predicted from a regression on the growth in Chinese imports by other South American countries. This estimation can be thought of as the first stage of an Instrumental Variable regression and it is aimed at isolating the variation in the data that is due to structural changes occurring in China and not due to events taking place in Brazil. Figure 3.6 in the appendix shows that the correlation between the two growth series was close to 90 percent. Both the observed and predicted annualized growth in Brazilian imports from China range between 10 percent and 45 percent depending on the sector. However impressive these growth rates may appear, we cannot gauge the magnitude of the shock from workers' point of view without looking at the levels. In figure 1.6 I divide the change in real imports between 2000 and 2010 by the number of workers initially employed in each sector. I observe large contrasts even within Manufacturing, where the change in imports per worker across sectors is up to two orders of magnitude apart. This exercise highlights the large variation in trade exposure by sector and complements the spatial variation shown in figure 1.2.

I use a static version of the model to translate the predicted sectoral changes in imports from China to Brazil between 2000 and 2010 into changes in Chinese productivity. In figure 1.7 I show that the model rationalizes the increase in imports by predicting large productivity gains in China across all sectors. However, the correlation is not perfect as other general equilibrium mechanisms come into play. I feed these annualized productivity gains into the second version of the model over the same period. In the structural analysis below I refer to this increase in Chinese productivity as the Chinese shock. Finally, I assume that Chinese productivity stops growing after ten periods and stays constant in all future periods. All other parameters are assumed to be constant throughout

Figure 1.6: Actual and predicted change in Brazilian imports from China per worker



Notes: Actual change in Brazilian imports from China per worker, between 2000 and 2010 (in 2010 constant US dollars). Brazilian workforce by industry is measured in 2000. Predicted change is computed from a regression of growth in Brazilian imports from China on growth in imports from China by other South American countries, as shown in figure 3.6. Data: BACI and Brazilian census.

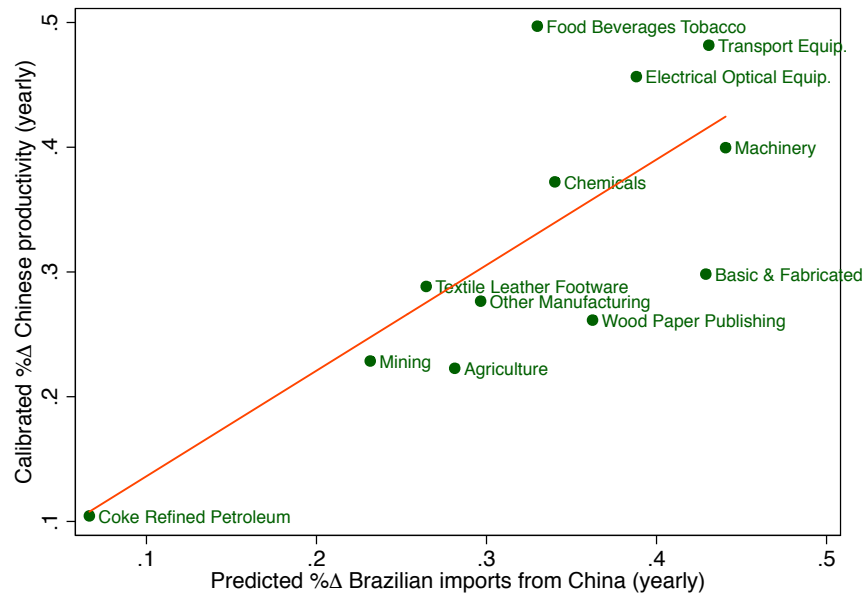
the analysis.

1.4.3 Main results

In this section I present the main results from the structural analysis. I put particular emphasis on three aspects of the results. First, unlike in the reduced-form analysis where all the measured effects were relative to the average region, the structural analysis enables me to talk about absolute as well as relative effects. Second, the model lets me go beyond the study of employment and nominal effects by providing a structure to think also about welfare effects. Third, my dual analysis at the State and metropolitan area level allows me to compare whether the measured effects depend on the level of spatial aggregation in the data.

In figure 1.8, I present the employment shares by aggregate sector predicted by the model under two parameterization. In the first one, I assume that productivity levels in

Figure 1.7: Predicted growth in Brazilian imports from China and growth in Chinese productivity implied by the model



Notes: Annualized predicted growth in real imports from China to Brazil and annualized change in Chinese productivity implied by the model. Correlation = 0.75.

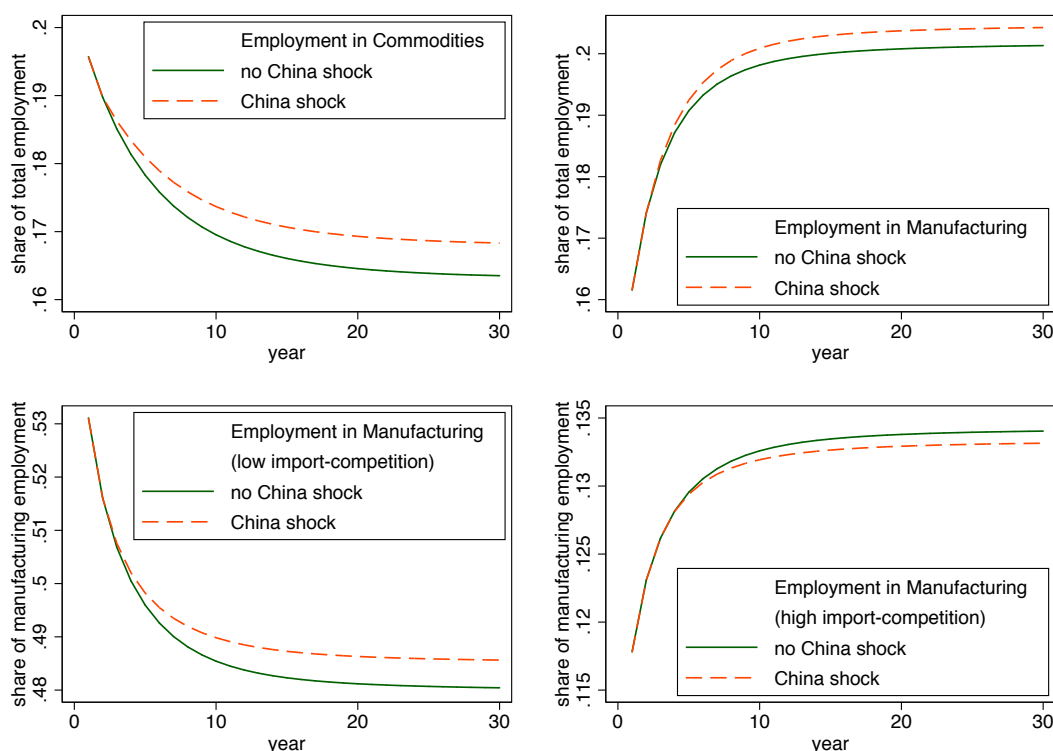
China do not change between 2000 and 2010. In the second one, Chinese sectoral productivities are calibrated such that the increase in Brazilian imports from China between 2000 and 2010 predicted by the model exactly matches the growth in imports predicted by the instrumentation strategy discussed earlier. The first important result is that from a national point of view the Brazilian economy was going through a long term shift in its production structure. Regardless of the China shock, the economy was moving away from the Primary sector and towards more Manufacturing and Services. It is in the context of this baseline scenario that I should analyse the impact of increased trade with China on Brazilian regions. The model suggests that relative to the baseline, employment in the Commodities sector shrank at a slower pace due to the China shock and its long term level is permanently higher than in the absence of the shock. By comparison, the reduced-form analysis suggested that employment in the regions specialized in export-oriented production grew more (or decreased less) than in the average region, this effect was driven primarily by employment in the Commodities sector. However, with the

differences-in-differences approach we could not tell whether the overall trend is upward or downward. With the structural analysis we can.

Second, perhaps surprisingly the model predicts that employment in the Manufacturing sector *grew* faster and to a permanently higher level due to the shock. A closer look at Manufacturing subsectors reveals that this is still consistent with the intuition that sectors more exposed to import competition from China suffered more. In particular, I focus on the three Manufacturing sectors most exposed and the three least exposed to import competition, as measured by the change in imports per worker shown in figure 1.6. Growth in Manufacturing employment is primarily driven by the subset of sectors relatively less exposed to Chinese import competition. On the other hand, the three Manufacturing sectors most exposed to the import shock continue growing but at a slower pace than in the absence of the shock.

While these employment shares were computed from the structural estimates based on data at the State level, the corresponding graph based on disaggregate data by metropolitan areas (see figure 3.7 in the appendix) shows very similar long-term trends. However, the rate of adjustment to the long-term levels varies with the level of aggregation. In the estimation at the State level, convergence to the long-term equilibrium is achieved much faster than in the estimation at the metropolitan area level. This is consistent with the following observation. In the model, labor markets are assumed to clear period by period and locally, however real wage differentials across locations take longer to disappear since spatial arbitrage only occurs via costly migration and with some delay. This implies that if there is any real wage differential within a sector and across two metropolitan areas in the same State, in the estimation on State data it is implicitly assumed that these differentials disappear instantaneously while in the estimation on more disaggregate data these differentials can last longer. However, in both cases the differentials have to disappear eventually unless the reallocation costs are prohibitively high. In other words, by estimating the model at a higher spatial scale, a researcher would underestimate the

Figure 1.8: Employment shares predicted by the model with and without China shock



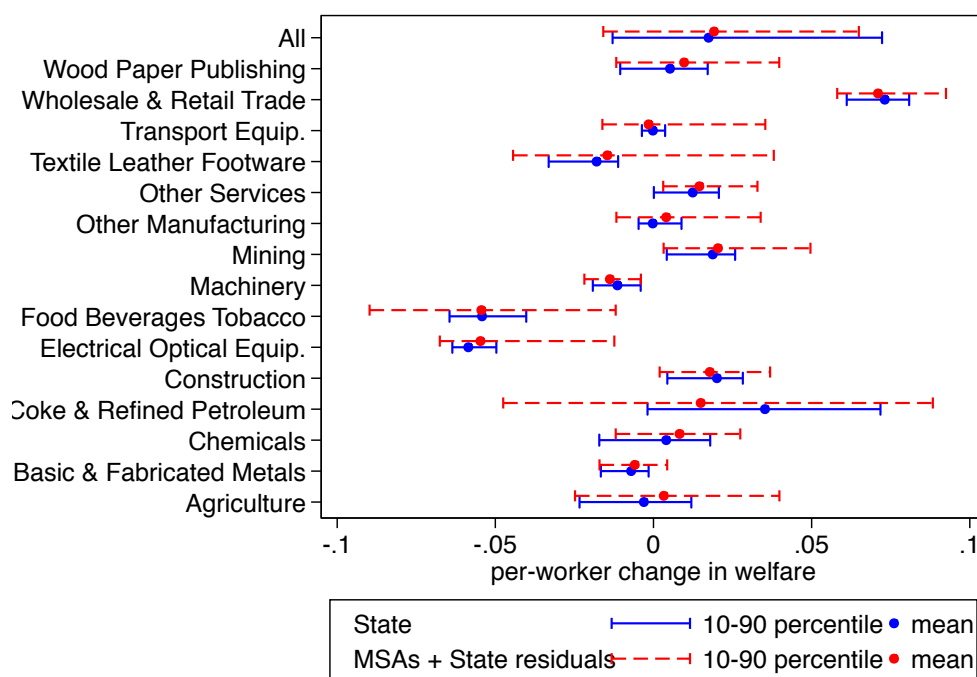
Notes: Evolution of sectoral employment shares predicted by the model using data at the State level. Commodity sector includes Agriculture and Mining. Low and high import competition manufacturing includes the three sectors with lowest and highest change in Brazilian imports from China per worker, as shown in figure 1.6.

length of the adjustment period relative to an estimation on data at the metropolitan area level.

Finally, I can study the welfare implications of increased trade with China. I take the perspective of a worker initially employed in a given region-industry pair and compare the discounted lifetime welfare she would enjoy with the shock to her welfare in the absence of the shock. Consistently with the model, her welfare is computed as a discounted sum of real wages and it also includes the option value of a costly relocation across regions, across industries, or both. In figure 1.9 I show the spatial and sectoral distribution of changes in workers' welfare predicted in the two exercises, i.e. using data at the State level and at the metropolitan area level.

The first observation is that on average, from a national perspective and across all

Figure 1.9: Lifetime welfare effects by initial sector



Notes:

sectors, Brazilian workers benefitted from increased trade with China. The model predicts that their welfare was between 1.7 and 1.9 percent higher than without the shock.

Second, judging from the sectoral mean changes in welfare, the impact was very different across sectors of initial employment. Workers in sectors where China had the highest productivity gains, as shown in figure 1.7, were among the ones who suffered the largest drop in welfare. Workers in the Non-Tradable sectors and in Mining were among the ones whose welfare increased the most.

The third and last observation is that within sectors there is substantial spatial heterogeneity in welfare effects. For each sector, I plot the the 10th and 90th percentile of the distribution of welfare changes across locations. The percentiles are computed using employment shares, so they represent percentiles of total population regardless of whether the analysis is at the level of States or metropolitan areas. While the mean effects are not too sensitive to the level of spatial aggregation in the data, the analysis at the State level vastly underestimates the amount of spatial heterogeneity in welfare effects found at a

more disaggregate level. Again, this confirms that by collapsing all distinct local labor markets within a State into a single one, we implicitly affect the scale at which the economic mechanisms in the model operate. For instance, local price indexes reflect local congestion. However, as in the example discussed in section 1.3, if a State consists of a declining metropolitan area and a distinct growing one, an analysis at the State level effectively underestimates both the decline in the cost of living in one location and its growth in the other one, which is the same as averaging them out. This explains my finding of less dispersed welfare outcomes when I estimate the model at the State level.

1.5 CONCLUSION

In this paper I estimate the effects of increased trade with China on Brazilian local labor markets. I show that there was substantial variation in local exposure to trade as measured by the observed change in Brazilian imports from and exports to China at the level of municipalities between 2000 and 2010. My units of analysis are commuting zones constructed from the bottom-up using commuting micro-data. I adopt reduced-form estimation strategies commonly used in the literature in order to compare my results to previous findings. My analysis at the regional level suggests that nominal earnings grew more in the exporting regions and less in the import-competing regions than in the average metropolitan area. Exits from the formal sector and in-migration also seem to depend on the sign and degree of local exposure to Chinese trade. However, I also show that these reduced-form approaches lead to estimates that are potentially biased as they do not take into account that regions are linked to one another by flows of factors and goods. I use individual-level migration histories to estimate how this localized shock diffused over time to other locations via migration. I find that after 10 years up to 30 percent of the shock is absorbed via migration, as measured by the correlation between average employment status and monthly wage of the cohort of initial residents and those of subsequent

residents.

I complement the reduced-form analysis at the regional and individual level with a structural analysis based on a model by [Caliendo *et al.* \(2015\)](#). This is the first estimation of their model using data at the metropolitan area level. The model allows me to endogenize the flows of factors and goods across domestic locations and it also lets me study the welfare effects of increased trade with China, which was not possible in the reduced-form analysis due to the lack of detailed price data. I calibrate the shock as an increase in Chinese sectoral productivities that is consistent with the observed increase in Brazilian imports from China and by other South American countries. I find that in the absence of the Chinese shock the Brazilian economy would have gone through a structural shift. That is, the employment share of the Commodities sector would have dropped and that of Manufacturing and Services would have increased. Relative to these baseline trends, the employment effect of increased trade with China at the national level was a slower reduction in the employment share of the Commodities sector and a slower growth in the Manufacturing subsectors that were relatively more exposed to Chinese import competition. Finally, my structural analysis suggests that while the average Brazilian worker benefitted from this shock, the welfare effects were very heterogeneous across sectors and across locations. I find that this heterogeneity is vastly underestimated if instead of using data at the level of metropolitan areas I use data aggregated by States and I explain why the choice of spatial units affects these results.

Chapter 2

Agglomeration: A Long-Run Panel Data Approach

2.1 INTRODUCTION

What are the key factors driving city growth over the long term?¹ One of the leading answers to this question, dating back to [Marshall \(1890\)](#), is that firms may benefit from proximity to one another through agglomeration economies. While compelling, this explanation raises further questions about the nature of these agglomeration economies. Do firms primarily benefit from proximity to other firms in the same industry, or, as suggested by [Jacobs \(1969\)](#), is proximity to other related industries more important? How do these forces vary across industries? What role does city size play in industry growth? How can we separate all of these features from the fixed locational advantages of cities? These are important questions for our understanding of cities. Their answers also have implications for the design of place-based policies, which can top \$80 billion per year in the U.S. and are also widely used in other countries.²

Not surprisingly, there is a large body of existing research exploring the nature of agglomeration economies. This study builds on two important strands of this literature.³ One approach uses long-differences in the growth of city-industries over time and relates them to rough measures of initial conditions in a city, such as an industry's share of city employment or the Herfindahl index over major city-industries ([Glaeser *et al.* \(1992\)](#), [Henderson *et al.* \(1995\)](#)). The main concern with this line of research is that it ig-

¹This chapter is joint work with Walker Hanlon (UCLA and NBER). We thank David Albouy, Pierre-Philippe Combes, Dora Costa, Don Davis, Jonathan Dingel, Gilles Duranton, Glenn Ellison, Ben Faber, Pablo Fajgelbaum, Edward Glaeser, Laurent Gobillon, Richard Hornbeck, Matt Kahn, Petra Moser, Alex Whalley and seminar participants at Columbia, UCLA, Harvard, UC Merced, UC San Diego, the NBER Innovation group, the NBER Urban Economics group, the CURE conference at Brown University, and the Urban Economics Association Annual Conference, for helpful comments and suggestions. Reed Douglas provided excellent research assistance. Funding for this project was provided by a grant from UCLA's Ziman Center for Real Estate.

²The *New York Times* has constructed a database of incentives awarded by cities, counties and states to attract companies to locate in their area. The database is available at <http://www.nytimes.com/interactive/2012/12/01/us/government-incentives.html>.

³There are several other strands of the agglomeration literature which are less directly related to this paper. Other alternative approaches use individual-level wage data ([Glaeser & Mare \(2001\)](#), [Combes *et al.* \(2008\)](#), [Combes *et al.* \(2011\)](#)) or firm-level data ([Dumais *et al.* \(2002\)](#), [Rosenthal & Strange \(2003\)](#), [Combes *et al.* \(2012\)](#)) to investigate the effects of city size. See [Rosenthal & Strange \(2004\)](#) and [Combes & Gobillon \(2015\)](#) for reviews of this literature.

nores much of the richness and heterogeneity that are likely to characterize agglomeration economies. A more recent approach allows for a richer set of inter-industry relationships using connection matrices based on input-output flows, labor force similarity, or technology spillovers. These connections are then compared to a cross-section of industry locations (Rosenthal & Strange (2001), Ellison *et al.* (2010), Faggio *et al.* (Forthcoming)).⁴ A limitation of this type of static exercise is that it is more difficult to control for locational fundamentals in cross-sectional regressions.

Our approach builds on these previous studies, but also seeks to address some of the remaining issues facing the literature. Specifically, this study contributes to the existing literature in five ways. First, while this is primarily an empirical paper, we begin by introducing a new dynamic spatial equilibrium model of city-industry growth. This model incorporates a rich set of within- and cross-industry spillover effects, which allows us to ground our study of these agglomeration forces in a theoretically-consistent framework. Recent work has highlighted the need for theoretical foundations in this literature.⁵

Second, motivated by the theory, we introduce a panel-data econometric approach for estimating the magnitude of agglomeration forces.⁶ The key feature of our approach is that we are able to estimate the importance of dynamic agglomeration forces related to industry scale, cross-industry connections, and city-size in a unified framework, while dealing with fixed locational fundamentals and time-varying industry-specific shocks. Previous research has examined these elements separately, but we are not aware of existing work that studies all of these effects in a unified way. In addition, the use of panel data offers some well-known advantages relative to the cross-sectional or long-difference methods used in most existing work. However, applying this approach to study agglom-

⁴These studies are part of a broader literature looking at the impact of inter-industry connections, particularly through input-output linkages, that includes work by Amiti & Cameron (2007) and Lopez & Sudekum (2009).

⁵See the handbook chapter by Combes & Gobillon (2015).

⁶Our panel data approach builds on previous work by Henderson (1997) and Dumais *et al.* (1997). See also Combes (2000) and Dekle (2002). A panel data approach is also used in a recent working paper by Lee (2015) which uses data on U.S. manufacturing industries from 1880-1990 to study static agglomeration forces.

eration economies requires overcoming challenges related to identification and correlated errors. Our study makes progress in this direction, allowing us to address some of the identification concerns present in previous work. The approach that we develop can potentially be applied in a wide range of settings in which consistent panels of city-industry employment data can be constructed.

Third, to implement our approach, we construct a rich dataset describing the evolution of city-industry employment over six decades. The availability of detailed long-run city-industry data has been a major impediment to previous work on agglomeration economies. The database constructed in this study helps address this deficiency.⁷ These new data, which we digitized from original sources, cover 31 of the largest English cities (based on 1851 population) for the period 1851-1911. This empirical setting offers several important advantages. One advantage is the very limited level of government regulation and interference in the British economy during this period due to the strong free-market ideology that dominated British policymaking and the small size of the central government.⁸ A second important advantage is that we are able to study agglomeration using consistent data over many decades. Studying agglomeration over a long time period is desirable because the time needed to build new housing, factories, and infrastructure means that it may take years for cities to respond to changes in local productivity levels. Our data are also quite detailed; they come from a full census and cover nearly the entire private sector economy, including manufacturing, transportation, retail, and services. A third advantage is that we are able to study a long-established urban system. This con-

⁷Recently, other databases of this type have been developed using data from the U.S. County Business Patterns by [Duranton *et al.* \(2014\)](#) and from the U.S. Census of Manufacturers by [Lee \(2015\)](#) and others.

⁸This contrasts with modern settings, where the list of confounding factors includes place-based government policies, local land-use regulations such as zoning, environmental policies that vary across locations, local tax incentives, variation in the local burden of national taxation, as well as many other types of regulation. These factors can also affect city growth, making it more difficult to identify and quantify the role of agglomeration forces. To cite some examples, [Kline & Moretti \(2013\)](#) describe the impact of place-based government policies in the U.S. The role of local land use regulations is highlighted by [Gyourko *et al.* \(2008\)](#). Local environmental policies are studied by [Henderson \(1996\)](#) and [Chay & Greenstone \(2005\)](#), among others. [Greenstone & Moretti \(2003\)](#) describe the impact of local tax incentives, while [Albouy \(2009\)](#) describes how federal tax incentives distort urban growth.

trasts with the U.S., where the open western frontier meant that the U.S. city system was in transition until the middle of the 20th century.⁹ Our setting was also characterized by a relatively open economy with high levels of migration into and between cities.¹⁰

Fourth, we provide new results on the strength of different types of agglomeration and congestion forces for one empirical setting. We find that (1) cross-industry effects were important, and occurred largely through the presence of local suppliers and occupationally similar labor pools, (2) the net effect of within-industry agglomeration forces was generally negative, and (3) city size had a clear negative relationship to city growth. The presence of local buyers appears to have had little positive influence on city-industry growth. We provide a variety of tests examining the robustness of these results. For example, we show that our main results are robust to dropping particular cities or particular industries. They are also robust to using an alternative set of matrices measuring cross-industry connections, alternative functional forms for modeling spillovers, or alternative industry definitions. We also show that incorporating cross-city effects, such as market potential or cross-city industry spillovers, has little impact on our results.

Fifth, we introduce a novel approach for measuring the combined strength of the many cross-industry agglomeration forces represented in our model. This is valuable because it provides a convenient way to assess the aggregate strength of these effects and may be useful for studying how these effects vary in different circumstances. Our results suggest that a lower-bound estimate of the agglomeration forces captured by our empirical model are equivalent to a decadal city-size divergence rate of 1.6-2.3%. To our knowledge this is the first paper to show how to measure the combined strength of these many cross-industry connections.

It is important to understand at the outset that the goal of this paper is to assess the role of agglomeration economies in driving city employment growth in different industries,

⁹See [Desmet & Rappaport \(Forthcoming\)](#). In contrast, [Dittmar \(2011\)](#) finds that Zipf's Law emerged in European cities between 1500 and 1800, well before the beginning of our study period.

¹⁰See, e.g., [Baines \(1994\)](#) and [Long & Ferrie \(2004\)](#).

and thereby contributing to overall city growth. Because our interest is in city growth, our analysis focuses specifically on employment as the outcome variable of interest. This is the natural object for our analysis, and one of the few types of data that can be observed at a local level, for many locations, over long time periods.¹¹ While the contribution of agglomeration economies to employment growth is generated through improved productivity, there is not necessarily a one-to-one mapping between productivity and employment growth. For example, under certain circumstances productivity improvements may reduce employment growth. Thus, our results should not be interpreted as providing a full description of the productivity effects of agglomeration economies.

It is also important to note that this study focuses on dynamic agglomeration economies, i.e., the influence of the current level of economic activity on future growth. This approach is motivated by the endogenous growth literature, and in particular the work of [Lucas \(1988\)](#), who emphasized the important role that localized learning in cities is likely to play in generating sustained economic growth. In some sense our exercise can be thought of as a step towards identifying the patterns that characterize endogenous growth at the urban level. This approach contrasts with work studying static agglomeration effects, where the level of employment or output in one sector influences the level in another sector. While static agglomeration effects are worthy of study, ultimately they cannot provide a theory of sustained urban growth.¹²

This paper analyzes agglomeration patterns across sectors spanning the entire private-sector economy in all of the largest urban centers in England for a period of sixty years. This broad approach allows us to estimate general patterns and to assess their importance for long-run city development. An alternative strand of work on agglomeration economies focuses on overcoming identification issues by comparing outcomes in sim-

¹¹Other types of data, such as wages and rents, are more difficult to obtain in a consistent way at the local level over long periods.

¹²Some discussion of static vs. dynamic agglomeration forces is provided in [Combes & Gobillon \(2015\)](#). [Lee \(2015\)](#) provides a recent example of a study focusing on static agglomeration forces. He finds that static localized inter-industry spillovers were small and declining in the U.S. across the 20th century. This suggests that static agglomeration forces are unlikely to be behind the growth of cities during this period.

ilar locations, where some locations receive a plausibly exogenous shock to the level of local economic activity (e.g., [Greenstone *et al.* \(2010\)](#) and [Kline & Moretti \(2013\)](#)). This approach has the advantage of more cleanly identifying the causal impact of changes in local economic activity, but it may also be less generalizable and more difficult to apply to policy analysis. Thus, we view our broader approach, which follows the work of [Glaeser *et al.* \(1992\)](#), [Henderson *et al.* \(1995\)](#)), and more recently [Ellison *et al.* \(2010\)](#), as complementary to studies that improve identification by focusing on specific shocks to local economic activity.

The next section presents our theoretical framework while Section [2.3](#) describes the data. The empirical approach is discussed in Section [2.4](#). Section [2.5](#) presents the main results, while Section [2.6](#) examines the impact of city size and shows how this can be used to calculate the aggregate strength of the agglomeration forces in our model. Section [2.7](#) concludes.

2.2 THEORY

While this paper is primarily empirical, a theoretical model is useful in disciplining the empirical specification. Grounding our analysis in theory can also help us interpret the results while being transparent about potential concerns.

The model is dynamic in discrete time. Technology advances over time as a result of two forces. First, firms undertake R&D in order to improve their productivity. Second, some of the new innovations produced by R&D undertaken by one firm spillover to affect other local firms. These spillovers can occur both within and across industries and the extent of the spillovers depends on a matrix of parameters reflecting the strength of within and inter-industry connections. These spillovers are external to firms, so they will not influence the static allocation of economic activity.

At the end of each period, technology diffuses across firms in the same city and indus-

try. This approach, which follows [Desmet & Rossi-Hansberg \(2014\)](#), substantially simplifies the dynamic elements of the model because firm R&D decisions will only affect firm profits in the current period. By simplifying the dynamics in this way, we are able to build a tractable model with a rich set of inter-industry connections.

As is standard in urban theories, we assume that goods are freely traded across locations and workers are free to move between cities. To keep things simple, our baseline model omits some additional features, such as savings and capital investment, or intermediate inputs, that one might want to consider. In the Appendix, we explore the impact of adding capital or intermediate goods.¹³

We begin by solving the allocation of employment across space in a particular period. We then consider how the allocation in one period affects the evolution of technology and, thus, the allocation of employment in the next period, through knowledge spillovers. Most of the interesting features of the model are on the producer's side, but we begin with a very brief introduction of the consumers.

2.2.1 Consumption

The model is populated by two types of agents, workers and landlords. There is a continuum of workers in the model, each endowed with one unit of labor. Workers have the option of paying a fixed cost, in terms of labor, in order to become entrepreneurs and open up their own firm. The utility function for both workers and landlords is, $U = \sum_{t=0}^{\infty} u_t e^{-\rho t}$ where u_t is utility in period t . There is no saving, so utility is maximized period-by-period.¹⁴ Utility in any period depends on consumption of real estate h_{ct} and a composite of goods G_{ct} according to a Cobb-Douglas utility function:

¹³The inclusion of these elements does not change the basic estimating equation that we obtain as long as we maintain the assumption of free mobility across locations, though it can change the interpretation of the parameter estimates.

¹⁴Adding savings would complicate the model, but as long as capital is mobile across locations it will not alter our basic estimating equation, nor will it influence our empirical results, which are derived from a comparison across locations within a country.

$$u_t = h_{ct}^\nu G_{ct}^{1-\nu} \quad (2.1)$$

where $\nu \in (0,1)$. There are i types of goods available, each produced by a separate industry, and consumers have CES preferences over these goods, so,

$$G_{ct} = \left(\sum_i \gamma_i x_{ict}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}$$

where x_{ict} is consumption of type i goods by a consumer in city c , σ is the elasticity of substitution across goods and $\gamma_i > 0$ is a demand shifter for industry i . The corresponding price index, P_t takes the standard form, with the price of each type of good denoted by p_{it} . Note that, with free trade, goods prices are not indexed by c . The index of goods prices is normalized to $P_t = 1$. The price of housing is denoted by q_{ct} . Consumers maximize their utility subject to their budget constraint. This utility maximization problem yields the expected demand equations for goods and real estate.

Workers have access to a time-varying outside option utility v_t^* . We can think of this as the utility offered by remaining in the rural sector or immigrating to another country. In equilibrium, this implies that the indirect utility function of workers must satisfy,

$$V_{ct} = \ln(w_{ct}) - \nu \ln(q_{ct}) = v_t^* . \quad (2.2)$$

Landlords receive income from land and other local resources. To keep things simple, we think of these landlords as living outside of the cities we study.

2.2.2 Production

Workers can decide to become entrepreneurs by paying a fixed cost F , denominated in units of labor, to open a firm. The measure of firms in a city-industry is denoted by n_{ict} . We think of firms in a city as being started by workers from that city in the previous

period, so if they enter in industry i they begin with the initial technology level available in that industry in that city, denoted \bar{a}_{icft} . Firms then invest in R&D to obtain a new technology level, a_{icft} , which is used in production.

Firms compete on perfectly competitive input and output markets. The production function for firm f in industry i and city c is,

$$y_{icft} = a_{icft} L_{icft}^{\alpha_1} H_{icft}^{\alpha_2} R_{icft}^{\beta} E_{icft}^{1-\alpha_1-\alpha_2-\beta}, \quad (2.3)$$

where a_{icft} is technology, L_{icft} is labor input, R_{icft} is the resource input, H_{icft} is real estate input, E_{icft} is entrepreneurial effort, $\alpha_1 + \alpha_2 + \beta < 1$, and $\alpha_1, \alpha_2, \beta > 0$.

Entrepreneurial effort is supplied by workers who choose to open a firm. Each entrepreneur has access to only one unit of entrepreneurial effort, so in equilibrium $E_{icft} = 1$ for all firms. This reflects a span-of-control limitation for firm owners.¹⁵ This span-of-control limitation plays an important role in the model; by introducing decreasing returns to scale at the firm level it pins down firm size. As we will see, this implies that growth in city-industry employment is driven by growth in the number of firms.

Labor is the only production input that is mobile across locations.¹⁶ Including real estate in the production function is not central to the model but is done for completeness.

2.2.3 Land and natural endowments

Locational fundamentals play a central role in the debate over the determinants of city size, so it is important that they be incorporated into the theory (see, e.g., [Davis & Weinstein \(2002\)](#)). In our model, locational fundamentals are represented by fixed industry-

¹⁵Note that entrepreneurs are not required to trade off entrepreneurial effort against labor effort. Instead, all workers have access to one unit of each type of input, but entrepreneurial effort can only be used by workers that choose to open a firm.

¹⁶Adding additional mobile inputs, such as capital, would not substantially affect the estimating equation that we obtain.

specific city resource endowments, \bar{R}_{ic} .¹⁷ In equilibrium, the markets for local resources clear, so $\int_{f=0}^{n_{ict}} R_{icft} df = \bar{R}_{ic}$. Resources play an important role in the model; by introducing decreasing returns at the city-industry level, they allow firms in the same industry to be active in many locations with different technology levels, even when trade is free, labor is mobile, and firms are perfectly competitive. They are also important in the context of the empirical analysis, because they make the impact of locational fundamentals in the estimation strategy explicit.

Real estate, which is used by both workers and firms, represents a congestion force in our theory. We model the price of real estate as an increasing function of the number of workers in a city and the amount of land used by producers:

$$q_{ct} = f \left(L_{ct}, \sum_i \sum_f H_{icft} \right). \quad (2.4)$$

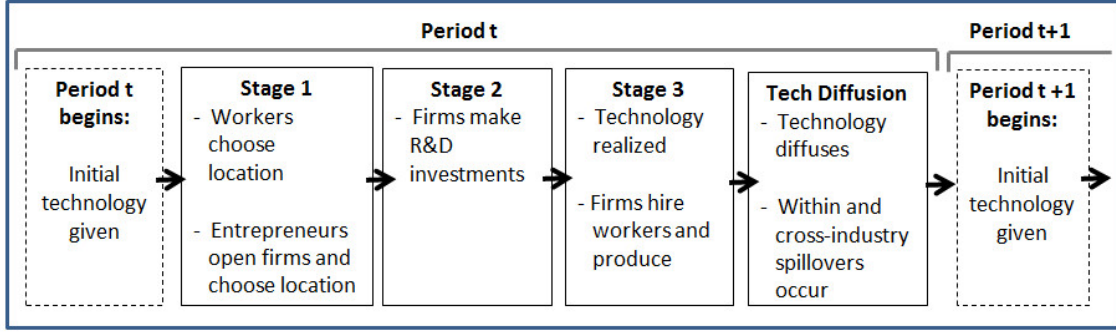
For our purposes, it is not necessary that we take a stand on the particular functional form of this relationship.

2.2.4 Timing

Figure 2.1 describes the timing in the model. At the beginning of each period, firms in the same city-industry share a common and observable technology level denoted \bar{a}_{icft} . Given these, workers choose where to locate and whether to pay a fixed cost to become an entrepreneur and open a firm. After workers have moved and firms have opened, firms then choose a level of R&D investment and realize a new technology level a_{icft} . Once technology is realized, firms choose how many workers and other inputs to hire and they produce and sell their outputs. At the end of the period, technology diffusion and technology spillovers occur, leading each firm to attain a new technology level \bar{a}_{icft+1} . The static portion of the model is solved by starting at Stage 3 and solving backwards.

¹⁷This approach follows Jones (1975) and has recently been used to study the regional effects of international trade by ? and Dix-Carneiro & Kovak (2015).

Figure 2.1: Model timing



2.2.5 Production: Stage 3

At the beginning of stage three, the number of workers in a city, L_{ct} , the number of firms in a city-industry, n_{ict} , and the technology level available to each firm a_{icft} have been determined. Given these, firms maximize profits by solving,

$$\max_{L_{icft}, H_{icft}, R_{icft}} p_{it} a_{icft} L_{icft}^{\alpha_1} H_{icft}^{\alpha_2} R_{icft}^{\beta} - w_{ct} L_{icft} - q_{ct} H_{icft} - r_{ict} R_{icft}$$

where w_{ct} is the wage and r_{ict} is the price of local resources. Since entrepreneurs will employ all of the entrepreneurial effort available to them, E_{ict} is not included in this optimization problem. Using the first order conditions for this expression, gross profits – which are the returns to entrepreneurial effort excluding fixed costs of entry and R&D expenditures – are:

$$\pi_{icft} = \left(\frac{w_{ct}^{\alpha_1} q_{ct}^{\alpha_2} r_{ict}^{\beta}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \beta^{\beta}} \right)^{\frac{-1}{1-\alpha_1-\alpha_2-\beta}} p_{it}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} a_{icft}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} (1 - \alpha_1 - \alpha_2 - \beta) \quad (2.5)$$

Local resource market clearing allows us to solve for the rental rate:

$$r_{ict} = \bar{R}_{ic}^{\frac{-(1-\alpha_1-\alpha_2-\beta)}{1-\alpha_1-\alpha_2}} \left(\frac{w_{ct}^{\alpha_1} q_{ct}^{\alpha_2}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \beta^{1-\alpha_1-\alpha_2}} \right)^{\frac{-1}{1-\alpha_1-\alpha_2}} p_{it}^{\frac{1}{1-\alpha_1-\alpha_2}} \left(\int_{f=0}^{n_{ict}} a_{icft}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} df \right)^{\frac{1-\alpha_1-\alpha_2-\beta}{1-\alpha_1-\alpha_2}} \quad (2.6)$$

2.2.6 Producers: Stage 2

At the beginning of stage two, workers and firms have already made their location decisions and firms have access to an initial technology level \bar{a}_{icft} . Given these, firms must choose how much to invest in R&D to increase their productivity in order to maximize profits. In doing so, they take into account the production decisions that we solved for above.

When firms conduct R&D, they are choosing a technology multiplier $\phi_{icft} \geq 0$ that increases their initial technology level according to,

$$a_{icft} = (1 + \phi_{icft})^\delta \bar{a}_{icft}, \quad (2.7)$$

at a cost $w_{ct} C \phi_{icft}$, where C is a parameter that determines the labor cost of innovation.¹⁸

We assume that $\delta < 1 - \alpha_1 - \alpha_2 - \beta$ so that the firm's profit function is concave in the R&D investment level.

Firms choose the innovation investment that maximizes gross profits less R&D expenses,

$$\max_{\phi_{icft}} \left(1 + \phi_{icft}\right)^{\frac{\delta}{1-\alpha_1-\alpha_2-\beta}} \bar{a}_{icft}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} p_{it}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} \left(\frac{w_{ct}^{\alpha_1} q_{ct}^{\alpha_2} r_{icft}^\beta}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \beta^\beta}\right)^{\frac{-1}{1-\alpha_1-\alpha_2-\beta}} (1 - \alpha_1 - \alpha_2 - \beta) - w_{ct} C \phi_{icft}$$

subject to $\phi_{icft} \geq 0$. For now, assume that there is an interior solution to this problem, so that $\phi_{icft} > 0$. In this case, the first order conditions for the firm's problem can be used to obtain the following expression for the firm's R&D decision, where the resource rent is substituted out using Eq. 2.6:

¹⁸While we do not allow the R&D cost to vary by industry here, allowing an industry-specific cost parameter would not fundamentally alter our results.

$$\begin{aligned}
(1 + \phi_{icft}) = & \left[\left(\frac{\delta}{C} \right)^{1-\alpha_1-\alpha_2-\beta} \bar{a}_{icft} \bar{R}_{ic}^{\frac{\beta(1-\alpha_1-\alpha_2-\beta)}{1-\alpha_1-\alpha_2}} w_{ct}^{\frac{(\alpha_2-1)(1-\alpha_1-\alpha_2-\beta)}{1-\alpha_1-\alpha_2}} p_{it}^{\frac{1-\alpha_1-\alpha_2-\beta}{1-\alpha_1-\alpha_2}} \right. \\
& \left. \left(\frac{q_{ct}^{\alpha_2}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2}} \right)^{\frac{-(1-\alpha_1-\alpha_2-\beta)}{1-\alpha_1-\alpha_2}} \left(\int_{f=0}^{n_{ict}} a_{icft}^{\frac{1}{1-\alpha_1-\alpha_2-\beta}} \right)^{\frac{-\beta(1-\alpha_1-\alpha_2-\beta)}{1-\alpha_1-\alpha_2}} \right]^{\frac{1}{1-\alpha_1-\alpha_2-\beta-\delta}} \quad (2.8)
\end{aligned}$$

Recalling that $1 - \delta - \alpha_1 - \alpha_2 - \beta > 0$, this equation tells us that for an individual firm the optimal level of innovation is increasing in the firm's initial technology level and the city-industry resource endowment. The level of innovation is decreasing in the cost of R&D, the wage level, and the amount of competition the firm faces for local resources, represented by the integral over the technology level of all other firms in the city-industry.

Suppose for now that all firms in an industry start with the same initial technology level \bar{a}_{icft} . Later, we will see that this is the case given how the technology diffusion process is modeled.¹⁹ In this case, firms in the industry will face the same R&D optimization problem, which implies that they will all choose the same R&D investment level, which we label ϕ_{ict}^* . Firms will be aware of the R&D decisions made by other firms and will take this into account when making their own decisions. The R&D investment consistent with these expectations is found by substituting Eq. 2.7 into Eq. 2.8 and solving to obtain,

$$(1 + \phi_{ict}^*) = \left[\left(\frac{\delta}{C} \right)^{1-\alpha_1-\alpha_2} \bar{a}_{icft} \bar{R}_{ic}^{\beta} \left(\frac{q_{ct}^{\alpha_2}}{\alpha_1^{\alpha_1} \alpha_2^{\alpha_2}} \right)^{-1} p_{it} w_{ct}^{\alpha_2-1} n_{ict}^{-\beta} \right]^{\frac{1}{1-\alpha_1-\alpha_2-\delta}} \quad (2.9)$$

This expression tells us that the level of innovation by a firm is increasing in the firm's initial technology level. At the same time it is decreasing in the number of firms in the same city-industry, which implies more competition for fixed city-industry resources.

¹⁹To keep things simple, and because firm heterogeneity is not central to the exercise undertaken in this paper, we have decided not to include within-industry firm heterogeneity in the model. However, firm heterogeneity could potentially be incorporated in a more sophisticated version of the model.

2.2.7 Producers: Stage 1

Next, we consider the entry decisions of firms. Any worker can choose to become an entrepreneur by paying a fixed cost (in terms of labor) of F . Because there is a large supply of potential entrants, *ex post* profits will be driven to zero. Thus, in equilibrium $\pi_{icft} - w_{ct}C\phi_{icft} = w_{ct}F$. Using this zero profit condition together with Eqs. 2.5, 2.6, and 2.9, we solve for the number of firms in a city-industry:

$$n_{ict} = w_{ct}^{\frac{\alpha_2-1}{\beta}} \left(\frac{q_{ct}^{\alpha_2}}{\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}} \right)^{\frac{-1}{\beta}} p_{it}^{\frac{1}{\beta}} \bar{a}_{ict}^{\frac{1}{\beta}} \left(\frac{1 - \alpha_1 - \alpha_2 - \beta - \delta}{F - C} \right)^{\frac{1 - \alpha_1 - \alpha_2 - \delta}{\beta}} \bar{R}_{ic} \left(\frac{\delta}{C} \right)^{\frac{\delta}{\beta}} \quad (2.10)$$

This, together with the first-order conditions from the firm's problem in Stage 3, gives city-industry employment:

$$L_{ict} = \alpha_1 w_{ct}^{\frac{\alpha_2-1}{\beta}} \left(\frac{q_{ct}^{\alpha_2}}{\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}} \right)^{\frac{-1}{\beta}} p_{it}^{\frac{1}{\beta}} \bar{R}_{ic} \bar{a}_{ict}^{\frac{1}{\beta}} \left(\frac{\delta}{C} \right)^{\frac{\delta}{\beta}} \left(\frac{1 - \alpha_1 - \alpha_2 - \beta - \delta}{F - C} \right)^{\frac{1 - \alpha_1 - \alpha_2 - \beta - \delta}{\beta}} \quad (2.11)$$

Together, Eqs. 2.10 and 2.11 imply the following relationship between the number of firms and the number of workers in a city-industry,

$$L_{ict} = \left(\frac{\alpha_1(F - C)}{1 - \alpha_1 - \alpha_2 - \beta - \delta} \right) n_{ict}. \quad (2.12)$$

Eq. 2.12 shows that growth in city-industry employment is driven entirely by growth in the number of firms. Eq. 2.10 can also be used, together with Eq. 2.9 to solve for the equilibrium level of R&D in the industry:

$$(1 + \phi_{ict}^*) = \frac{\delta(F - C)}{C(1 - \alpha_1 - \alpha_2 - \beta - \delta)} \quad (2.13)$$

This expression shows that firms' R&D investments will depend only on model parameters, a useful feature that simplifies the results. So far we have solved the model

assuming that $\phi_{ict}^* > 0$. For this to hold, we need, $F > C \left(\frac{1-\alpha_1-\alpha_2-\beta}{\delta} \right)$.²⁰ For the remainder of this theory we assume that this condition is satisfied so that we have an interior solution to the firm's R&D optimization problem and R&D occurs in all industries.

2.2.8 Spillovers and technology diffusion

At the end of a period, after production and consumption have taken place, firms are able to copy technology from other firms in the same industry (diffusion). However, because all firms in a given city-industry end the period with the same technology level, the role of diffusion is simply to rule out strategic behavior. In addition, entrepreneurs may share ideas, and this recombination of ideas can increase their productivity (spillovers). Following Glaeser *et al.* (1992), we write the growth rate in technology at the city-industry level as,

$$\ln \left(\frac{\bar{a}_{ict}}{a_{ict-1}} \right) = S_{ict-1} + \epsilon_{ict} \quad (2.14)$$

where $S_{ict-1} \geq 1$ represent the amount of spillovers available to a city-industry in a period. This can include within-industry effects, cross-industry spillovers, as well as national industry technology growth or city-level aggregate spillovers.

We can use Eq. 2.14 to translate the growth in (unobservable) city-industry technology into the growth of (observable) city-industry employment. Using Eq. 2.2, Eq. 2.7, Eq. 2.11, and Eq. 2.14, we obtain,

$$\begin{aligned} \Delta \ln(L_{ict+1}) &= \left(\frac{1}{\beta} \right) S_{ict} + \left(\frac{1}{\beta} \right) \Delta \ln(p_{it}) + \left(\frac{\alpha_2 - 1}{\beta} \right) \Delta \ln(\bar{v}_t^*) \\ &+ \left(\frac{\nu(\alpha_2 - 1) - \alpha_2}{\beta} \right) \Delta \ln(q_{ct}) + \zeta + \tilde{\epsilon}_{ict} \end{aligned} \quad (2.15)$$

²⁰Note that the expression in parenthesis can be interpreted as the ratio of the gains from additional firms in an industry to the gains from improved technology in the industry. For industries where the inequality above doesn't hold, there will be no innovation.

where ζ is a constant function of model parameters. Note that by differencing we have eliminated the local resource endowment from this equation.

As a final step, we need to decide how to model the spillover term. Existing empirical evidence provides little guidance here, so we will opt for a fairly simple approach.²¹ We model the spillovers benefits to firms in industry i from R&D in local firms in industry k as a function of (1) the amount of new ideas produced in industry k , which is a function of the size of the technology advance made by each firm $(1 + \phi_{kct}^*)$, and the number of firms, n_{kct} , and (2) the usefulness of these ideas to firms in industry i , given by parameter τ_{ki} . Thus, there is a matrix of τ_{ki} parameters representing the usefulness of an idea from industry k to producers in industry i . The diagonal τ_{ii} terms reflects within-industry spillovers.²² Given this, we write the spillover function as,²³

$$S_{ict} = \sum_k \tau_{ki} \ln(n_{kct}(1 + \phi_{kct})) + \zeta_{it} + \psi_{ct}.$$

Using Eqs. 2.12 and 2.13, this can be rewritten as,

$$S_{ict} = \sum_k \tau_{ki} \ln(L_{kct}) + \zeta_{it} + \psi_{ct} + \Gamma$$

where Γ is a constant term. Combining this with Eq. 2.15 we obtain,

²¹In the empirical analysis we will investigate the robustness of our results to some reasonable alternative formulations.

²²The intuition behind the within-industry spillovers in this model is that, while all firms achieve the same new technology level after undertaking R&D, this new level need not be achieved in exactly the same way. As a result, it may be possible for firms to achieve further gains by observing the different types of technologies developed by their competitors. However, the potential gains from within-industry spillovers will depend on a number of factors, such as the willingness for firms in an industry to share ideas with their direct local competitors.

²³Here we are assuming that city-industry resource endowments are such that $n_{kct} \geq 1$. This assumption allows us to express the spillover term in a slightly simpler way, but is not central to our results. If we are worried that n_{kct} can fall below one then we would instead write this as $S_{ict} = \sum_k \tau_{ki} \ln(\max(n_{kct}(1 + \phi_{kct}), 0)) + \zeta_{it} + \psi_{ct}$.

$$\begin{aligned}
\ln(L_{ict+1}) - \ln(L_{ict}) &= \left(\frac{1}{\beta}\right) \left[\tau_{ii} \ln(L_{ict}) + \sum_{k \neq i} \tau_{ki} \ln(L_{kct}) \right. \\
&+ \left. \left[\ln(p_{it}) - \ln(p_{it-1}) \right] + \xi_{it} \right. \\
&- \left. \left[v(\alpha_2 - 1) - \alpha_2 \right] \left[\ln(q_{ct}) - \ln(q_{ct-1}) \right] + \psi_{ct} \right. \\
&+ \left. (\alpha_2 - 1) \left[\ln(\bar{v}_t^*) - \ln(\bar{v}_{t-1}^*) \right] \right] + \tilde{\Gamma} + \tilde{\epsilon}_{ict}
\end{aligned} \tag{2.16}$$

where the constant terms have been gathered into $\tilde{\Gamma}$. This equation expresses the change in log employment in industry i and location c in terms of (1) within-industry spillovers generated by employment in industry i , (2) cross-industry spillovers, (3) national industry-specific factors that affect industry i in all locations, (4) city-specific factors that affect all industries in a location, and (5) aggregate changes in the outside option of workers that affect all industries in all locations.

This expression for city-industry growth will motivate our empirical specification. One feature that is worth noting here is that we have two factors, city-level aggregate spillovers ψ_{ct} and city congestion costs q_{ct} , both of which vary at the city-year level. Empirically we will not be able to separate these positive and negative effects and so we will only be able to identify their net impact. Similarly, we cannot separate positive and negative effects that vary at the industry-year level.

Note that in the absence of spillovers, and with common technologies across locations, the city size distribution in this model will be determined by the distribution of local resource endowments. Once local technology spillovers are added, city sizes will be determined by a combination of the initial resource endowment and the evolving technology levels. This hybrid of locational fundamentals and increasing returns is consistent with some existing empirical results (e.g., [Davis & Weinstein \(2002\)](#) and [Bleakley & Lin \(2012\)](#)). Once spillovers are included, the dynamics of the system are complex and depend cru-

cially on the matrix of τ_{ki} parameters.²⁴ Estimating these parameters is the goal of our empirical exercise, which we turn to next.

While our model provides a theoretically-grounded estimation approach, we recognize that this is not the only potential set of agglomeration forces that can yield an estimation equation that matches the one that we will apply. There are at least two promising alternative theories that may yield similar expressions. One such theory could combine static inter-industry connections, such as pecuniary spillovers through intermediate-goods sales, with changing transport costs. A second alternative combines static agglomeration forces with a friction that results in a slow transition towards a static equilibrium. Our empirical exercises cannot make a sharp distinction between the mechanisms described in our framework and these alternatives, so they should not be interpreted as a direct test of the particular agglomeration mechanism described by the theory. Rather, our empirical results will provide evidence on the pattern of within and cross-industry agglomeration benefits and provide some evidence on the types of inter-industry connections that matter. Further work will be needed to unpack the specific mechanisms through which these inter-industry benefits occur.

2.3 DATA

The main database used in this study was constructed from more than a thousand pages of original British Census of Population summary reports.²⁵ The decennial Census data were collected by trained registrars during a relatively short time period, usually a few days in April of each census year. As part of the census, individuals were asked to state their occupation, but the reported occupations correspond more closely to industries

²⁴In addition, the dynamics are likely to depend crucially on city-size congestion forces, which are not fully modeled here. Because the primary goals of this paper are empirical, we leave a full exploration of these dynamics for future work.

²⁵This study uses the most updated version of this database (v2.0). These data and further documentation can be found at <http://www.econ.ucla.edu/whanlon/> under Research.

than to what we think of as occupations today.²⁶ A unique feature of this database is that the information is drawn from a full census. Virtually every person in the cities we study provided information on their occupation and all of these answers are reflected in the employment counts in our data.²⁷

The database includes 31 cities for which occupation data were reported in each year from 1851-1911, containing 28-34% of the English population over the period we study. The geographic extent of these cities changes over time as the cities grow, a feature that we view as desirable for the purposes of our study.²⁸ Appendix 3.4 provides a list of the cities included in the database, as well as a map showing the location of these cities in England. In general, our analysis industries cover the majority of the working population of the cities, with most of the remainder employed by the government or in agriculture.

The industries in the database span manufacturing, food processing, services and professionals, retail, transportation, construction, mining, and utilities. Because the occupational categories listed in the census reports varied over time, we combined multiple industries in order to construct consistent industry groupings over the study period. This process generates 26 consistent private sector occupation categories.²⁹ Of these, 23 can be matched to the connections matrices used in the analysis. Table 3.2 in Appendix 3.4 describes the industries included in the database.

²⁶Examples from 1851 include “Banker”, “Glass Manufacture” or “Cotton manufacture”. The database does include a few occupations that do not directly correspond to industries, such as “Labourer”, “Mechanic”, or “Gentleman”, but these are a relatively small share of the population. These categories are not included in the analysis. In 1921 the Census office renamed what had previously been called “occupation” to be “industry” and then introduced a new set of data reflecting occupation in the modern sense.

²⁷This contrasts with data based on census samples, which often covers 5% or 1% of the available data. We have experimented with data based on a census sample (from the U.S.) and found that, when cutting the data to the city-industry level, sampling error has a substantial effect on the consistency and robustness of the results.

²⁸Other studies in the same vein, such as Michaels *et al.* (2013), also use metropolitan boundaries that expand over time. The alternative is working with fixed geographic units. While that may be preferred for some types of work, given the growth that characterizes most of the cities in our sample, using fixed geographic units would mean either that the early observations would include a substantial portion of rural land surrounding the city, or that a substantial portion of city growth would not be part of our sample in the later years. Either of these options is undesirable.

²⁹Individual categories in the years were combined into industry groups based on (1) the census’ occupation classes, and (2) the name of the occupation. Further details of this procedure are available at <http://www.econ.ucla.edu/whanlon/>.

A preliminary analysis, using the agglomeration measure from [Ellison & Glaeser \(1997\)](#), suggests that the agglomeration patterns observed in our data are similar to those documented in modern studies (details in Appendix 3.4, Tables 3.3-3.4). Britain's main manufacturing and export industries, such as Textiles, Metal & Machines, and Shipbuilding, show high levels of geographic agglomeration. Many non-traded services or retail industries, including Merchants, Agents, Etc., Construction, and Shopkeepers, Salesmen, Etc. show low levels of agglomeration. Overall, the median level of industry agglomeration is between 0.02 and 0.026, which is comparable to the levels reported for the modern U.S. economy by [Ellison & Glaeser \(1997\)](#) and somewhat larger than the levels reported for the modern British economy by [Faggio et al. \(Forthcoming\)](#).³⁰

This study also requires a set of matrices measuring the pattern of connections between industries. These measures should reflect the channels through which ideas may flow between industries. Existing literature provides some guidance here. [Marshall \(1890\)](#) suggested that firms may benefit from connections operating through input-output flows, the sharing of labor pools, or other types of technology spillovers. The use of input-output connections is supported by recent literature showing that firms share information with their customers or suppliers.³¹ To reflect this channel, we use an input-output table constructed by [Thomas \(1987\)](#) based on the 1907 British Census of Production (Britain's first industrial census).³² We construct two variables: $IOin_{ij}$, which gives the share of industry i 's intermediate inputs that are sourced from industry j , and $IOout_{ij}$ which gives the share of industry i 's sales of intermediate goods that are purchased by industry j .

³⁰Using industry data for 459 manufacturing industries at the four-digit level and 50 states, [Ellison & Glaeser \(1997\)](#) calculate a mean agglomeration index of 0.051 and a median of 0.026. For Britain, [Faggio et al. \(Forthcoming\)](#) calculate industry agglomeration using 94 3-digit manufacturing industries and 84 urban travel-to-work areas. They obtain a mean agglomeration index of 0.027 and a median of 0.009. [Kim \(1995\)](#) calculates an alternative measure of agglomeration for the U.S. during the late 19th and early 20th centuries, but given that he studies only manufacturing industries, and given the substantial differences between his industry definitions and our own, it is difficult to directly compare to his results.

³¹For example, [Javorcik \(2004\)](#) and [Kugler \(2006\)](#) provide evidence that the presence of foreign firms (FDI) affects the productivity of upstream and downstream domestic firms.

³²For robustness exercises, we have also collected an input-output table for 1841 constructed by [Horrell et al. \(1994\)](#) with 12 more aggregated industry categories. See Appendix 3.4 for more details.

One drawback of using these matrices is that they are for intermediate goods; they will not capture the pattern of capital goods flows.

Another channel for knowledge flow is the movement of workers, who may carry ideas between industries.³³ To reflect this channel, we construct two different measures of the similarity of the workforces used by different industries. The first measure is based on the demographic characteristics of workers (their age and gender) from the 1851 Census. These features had an important influence on the types of jobs a worker could hold during the period we study.³⁴ For any two industries, our demographic-based measure of labor force similarity, EMP_{ij} , is constructed by dividing workers in each industry into these four available bins (male/female and over20/under20) and calculating the correlation in shares across the industries. A second measure of labor-force similarity, based on the occupations found in each industry, is more similar to the measures used in previous studies. This measure is built using U.S. census data from 1880, which reports the occupational breakdown of employment by industry. We map the U.S. industry categories to the categories available in our analysis data. Then, for any two industries our occupation-based measure of labor force similarity, OCC_{ij} is the correlation in the vector of employment shares for each occupation.

Finally, it is worth noting that the transportation system connecting the cities in our database was relatively stable across the study period. All cities were connected by rail at the beginning of the study period, and many also had canal connections. Due in part to the stability of this system, as well as the importance of local resources such as coal, existing work suggests that changes in transport costs had little impact on the location of industry in Britain during this period (Crafts & Mulatu (2006)). At the same time, this well-developed transportation system contributed to the high levels of population mo-

³³Research by Poole (2013) and Balsvik (2011), using data from Brazil and Norway, respectively, has highlighted this channel of knowledge flow.

³⁴For example, textile industries employed substantial amounts of female and child labor, while metal and heavy machinery industry jobs were almost exclusively reserved for adult males.

bility that characterized this period.³⁵ This high level of labor mobility, together with the absence of regulatory restrictions on city growth, meant that city population and employment had the ability to respond to productivity improvements.

2.4 EMPIRICAL APPROACH

The starting point for our analysis is based on Equation 2.16, which represents the growth rate of a city-industry as a function of within and cross-industry agglomeration effects as well as time-varying city-specific and national industry-specific factors. Rewriting this as a regression equation we have,

$$\Delta \ln(L_{ict+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \sum_{k \neq i} \tilde{\tau}_{ki} \ln(L_{kct}) + \theta_{ct} + \chi_{it} + e_{ict} \quad (2.17)$$

where Δ is the first difference operator, $\tilde{\tau}_{ii}$ and $\tilde{\tau}_{ki}$ include $1/\beta$, θ_{ct} is a full set of city-year effects and χ_{it} is a full set of industry-year effects. The first term on the right hand side represents within-industry spillovers, while the second term represents cross-industry spillovers.³⁶

One issue with Equation 2.17 is that there are too many parameters for us to credibly estimate given the available data. In order to reduce the number of parameters, we need to put additional structure on the spillover terms. As discussed in the previous section, we follow recent literature in this area, particularly [Ellison *et al.* \(2010\)](#), by parameterizing the connections between industries using the available input-output and labor force

³⁵During this period the British population was “highly mobile” in the words of [Long & Ferrie \(2003\)](#). while [Baines \(1985\)](#) shows that population growth in cities was due in large part to the arrival of new migrants, coming both from the English countryside as well as Ireland, Scotland and Wales.

³⁶We purposely omitted the last term of Equation 2.16, $\Delta \ln(\bar{v}_t^*)$, because although it could be estimated as a year-specific constant, it would be collinear with both the (summation of) industry-year and city-year effects. Moreover, in any given year we also need to drop one of the city or industry dummies in order to avoid collinearity. In all specifications we chose to drop the industry-year dummies associated with the “General services” sector.

similarity matrices³⁷:

$$\tilde{\tau}_{ki} = \beta_1 IOin_{ki} + \beta_2 IOout_{ki} + \beta_3 EMP_{ki} + \beta_4 OCC_{ki} \quad \forall i, k$$

Substituting this into Eq. 2.17 we obtain:

$$\begin{aligned} \Delta \ln(L_{ict+1}) = & \tilde{\tau}_{ii} \ln(L_{ict}) + \beta_1 \sum_{k \neq i} IOin_{ki} \ln(L_{kct}) + \beta_2 \sum_{k \neq i} IOout_{ki} \ln(L_{kct}) \\ & + \beta_3 \sum_{k \neq i} EMP_{ki} \ln(L_{kct}) + \beta_4 \sum_{k \neq i} OCC_{ki} \ln(L_{kct}) + \theta_{ct} + \chi_{it} + e_{ict} \end{aligned} \quad (2.18)$$

Instead of a large number of parameters measuring spillovers across industries, Equation 2.18 now contains only four parameters multiplying four (weighted) summations of log employment. Summary statistics for the cross-industry spillover terms are available in Appendix Table 3.5 while the correlations between the cross-industry terms are available in Appendix Table 3.6.

There is a clear parallel between the specification in Equation 2.18 and the empirical approach used in the convergence literature (Barro & Sala-i Martin (1992)). A central debate in this literature has revolved around the inclusion of fixed effects for the cross-sectional units (see, e.g., Caselli *et al.* (1996)). In our context, the inclusion of such characteristics could help control for location and industry-specific factors that affect the growth rate of industry and are correlated with initial employment levels. However, the inclusion of city-industry fixed effects in Equation 2.18 will introduce a mechanical bias in our estimated coefficients (Hurwicz (1950), Nickell (1981)). This bias is a particular concern in a setting where the time-series is limited. Solutions to these issues have been offered by Arellano & Bond (1991), Blundell & Bond (1998), and others, yet these procedures can also

³⁷Adding an error term to this equation would imply heteroskedastic standard errors, a possibility that is accommodated by our econometric approach, but would not otherwise alter the basic estimation approach suggested by the theory.

generate biased results, as shown by [Hauk Jr. & Wacziarg \(2009\)](#). In a recent review, [Barro \(2012\)](#) uses data covering 40-plus years and argues (p. 20) that in this setting, “the most reliable estimates of convergence rates come from systems that exclude country fixed effects but include an array of X variables to mitigate the consequence of omitted variables.” Our approach essentially follows this advice, but with the additional advantage that we have two cross-sectional dimensions, which allows for the inclusion of flexible controls in the form of time-varying city and industry effects.

There are two issues to address at this point. First, there could be measurement error in L_{ict} . Since this variable appears both on the left and right hand side, this would mechanically generate an attenuation bias in our within-industry spillover estimates. Moreover, since L_{ict} is correlated with the other explanatory variables, such measurement error would also bias the remaining estimates. We deal with measurement error in L_{ict} on the right hand side by instrumenting it with lagged city-industry employment.³⁸ Under the assumption that the measurement error in any given city-industry pair is *iid* across cities and time, our instrument is $L_{ict}^{Inst} = L_{ict-1} \times g_{i-ct}$, where L_{ict-1} is the lag of L_{ict} and g_{i-ct} is the decennial growth rate in industry i computed using employment levels in all cities *except* city c , as in [Bartik \(1991\)](#).

Second, we are also concerned that there may be omitted variables that affect both the level of employment in industry j and the growth in employment in industry i . Such variables could potentially bias our estimated coefficients on both the cross-industry and (when $j = i$) the within-industry spillovers. For instance, if there is some factor not included in our model which causes growth in two industries i and $k \neq i$ in the same city, a naive estimation would impute such growth to the spillover effect from k to i , thus biasing the estimated spillover upward. Our lagged instrumentation approach can also help us deal with these concerns. Specifically, when using instruments with a one-decade lag to address endogeneity concerns the exclusion restriction is that there is not some

³⁸This approach is somewhat similar to the approach introduced by [Bartik \(1991\)](#) and has been suggested by [Combes et al. \(2011\)](#).

omitted variable that is correlated with employment in some industry k in period t and affects employment growth in industry i from period $t + 1$ to $t + 2$. Moreover, the omitted variable cannot affect growth in all industries in a location, else it would be captured by the city-year fixed effect, nor can it affect the growth rate of industry i in all cities.³⁹ Thus, while our approach does not allow us to rule out all possible confounding factors, it allows us to narrow the set of potential confounding forces relative to most previous work in this area. Now, for the cross-industry case, the summation terms in Equation 2.18 such as $\sum_{k \neq i} IOin_{ki} \ln(L_{kct})$ are instrumented with $\sum_{k \neq i} IOin_{ki} \ln(L_{kct}^{Inst})$, where L_{kct}^{Inst} is computed as described above.

The estimation is performed using OLS or, when using instruments, two-stage least squares. Correlated errors are a concern in these regressions. Specifically, we are concerned about serial correlation, which [Bertrand et al. \(2004\)](#) argue can be a serious concern in panel data regressions, though this is perhaps less of a concern for us given the relatively small time dimension in our data. A second concern is that industries within the same city are likely to have correlated errors. A third concern, highlighted by [Conley \(1999\)](#) and more recently by [Barrios et al. \(2012\)](#), is spatial correlation occurring across cities. Here the greatest concern is that error terms may be correlated within the same industry across cities (though the results presented in Appendix 2.3 suggest that cross-city effects are modest).

To deal with all of these concerns we use multi-dimensional clustered standard errors following work by [Cameron et al. \(2011\)](#) and [Thompson \(2011\)](#). We cluster by (1) city-industry, which allows for serial correlation; (2) city-year, which allows for correlated errors across industries in the same city and year; and (3) industry-year, which allows for spatial correlation across cities within the same industry and year. This method relies on asymptotic results based on the dimension with the fewest number of clusters. In our case this is $23 \text{ industries} \times 6 \text{ years} = 138$, which should be large enough to avoid serious

³⁹The results are not sensitive to the length of the lag used in the instrumentation. We have experimented with two- and three-decade lags and obtained essentially the same results.

small-sample concerns.

In order to conduct underidentification and weak-instrument tests while clustering standard errors in multiple dimensions, we have produced a new statistical package following the approach from [Kleibergen & Paap \(2006\)](#). This was necessary because existing statistical packages are unable to calculate these tests correctly when clustering by more than two dimensions. The procedure used to generate our new statistical package is described in Appendix 2.3. Our package, which we plan to make publicly available, can accommodate clustering across an arbitrary number of dimensions, which is likely to be useful for future researchers.

Finally, we may be concerned about how well our estimation procedure performs in a data set of the size available in this study. To assess this, we conduct a series of Monte Carlo simulations in which we construct 500 new data sets with a size and error structure based on the true data, but with known spillover parameter values. We then apply our estimation procedure to these simulated data in order to obtain a distribution of placebo coefficient estimates, which can then be compared to the estimates obtained using the true data. These simulations, which are described in more detail in Appendix 3.4, suggest that our estimation procedure performs well in datasets with a size and error structure similar to the true data.

To simplify the exposition, we will hereafter collectively refer to the set of regressors $\ln(L_{ict})$ for $i = 1 \dots I$ as the *within* variables. Similarly, with a small abuse of notation the term $\sum_{k \neq i} IOin_{ki} \ln(L_{kct})$ is referred to as *IOin*, and so on for *IOout*, *EMP*, and *OCC*. We collectively refer to the latter terms as the *between* regressors since they are the parametrized counterpart of the spillovers across industries.

2.5 MAIN RESULTS

Our main regression results are based on the specification described in Equation 2.18. The estimation strategy involves using four measures for the pattern of cross-industry spillovers: forward input-output linkages, backward input-output linkages, and two measures of labor force similarity. We begin our analysis in Table 2.1 by looking at results that include only one of these at a time. Columns 1-3 include only the forward input-output linkages; Column 1 presents OLS results; Column 2 presents results with lagged instrumentation on the within terms; and Column 3 uses lagged instrumentation for both the within and between terms. A similar pattern is used for backward input-output linkages in Columns 4-6, the demographic-based labor force similarity measure in Columns 7-9, and the occupation-based labor force similarity measure in Columns 10-12. All of these results include a full set of industry-specific within-industry terms, but these are not reported in Table 2.1 for space reasons.⁴⁰

These results show strong positive effects operating through forward input-output connections, suggesting that local suppliers play an important role in industry growth. The importance of local suppliers to industry growth is perhaps the clearest and most robust result emerging from our analysis. There is little evidence of positive effects operating through local buyers. The results do provide some evidence that the presence of other industries using similar labor pools may increase growth, particularly when using the more detailed OCC measure. A comparison across columns for each spillover measure shows that the IV results do not differ from the OLS results in a statistically significant way, suggesting that any measurement error or omitted variables concerns addressed by instruments are not generating substantial bias in the OLS results.

⁴⁰We do not report first-stage results for our instrumental variables regressions because these involve a very large number of first-stage regressions. Instead, for each specification we report the test statistics for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006) as well as the test statistic for weak instruments test based on the Kleibergen-Paap Wald statistic. It is clear from these statistics that weak instruments are not a substantial concern in these specifications.

Table 2.1: OLS and IV regressions including only one spillover path at a time

	(1)	(2)	(3)	(4)	(5)	(6)
IOin	0.0559*** (0.0152)	0.0458*** (0.0161)	0.0446*** (0.0158)			
IOout				-0.0084 (0.0108)	-0.0158 (0.0107)	-0.0192* (0.0107)
Observations	4,253	3,539	3,539	4,253	3,539	3,539
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
KP under id.		21.02	28.13		17.87	18.1
KP weak id.		51.94	61.7		43.72	42.34

	(7)	(8)	(9)	(10)	(11)	(12)
EMP	0.0017 (0.0016)	0.0029** (0.0015)	0.0018 (0.0014)			
OCC				0.0058* (0.0030)	0.0070** (0.0032)	0.0066** (0.0032)
Observations	4,253	3,539	3,539	4,253	3,539	3,539
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
KP under id.		19.22	20.29		15.72	16.9
KP weak id.		47.61	44.15		37.7	31.96

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heterogeneous *within* terms, city-time and industry-time effects are included in all regressions but not displayed. 2SLS regressions use lagged instruments. Note that the number of observations falls for the instrumented regressions because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*. “KP under id.” denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). “KP weak id.” denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic.

Table 2.2 considers all four channels simultaneously. Columns 1-3 present results in which we estimate a single coefficient on the within-industry terms. Columns 4-6 present results in which we estimate industry-specific within-industry effects. These heterogeneous within-industry coefficients, which are not reported in Table 2.2, will be explored later. Columns 1 and 4 presents OLS results. In Column 2 and 5 we instrument the within terms. In Column 3 and 6 we use instruments for both the within and between terms. The results are generally similar to those from Table 2.1; the presence of local suppliers or industries employing a similar labor force both appear to enhance city-industry growth. The presence of local buyers has no positive effect. In Columns 1-3, we can see that the within term is negative, suggesting that on average across all industries employ-

ment growth is slower in locations where initial industry employment is already large. Comparing the results in Columns 1-3 to those in Columns 4-6 shows how important it is to allow for variation in the within-industry effects, something which has often been impossible in previous studies.

Based on the results from Column 6 of Table 2.2, our preferred specification, a one standard deviation increase (2.45) in the presence of local suppliers (the *IOin* channel) increases city-industry growth by 14.4%. Turning to the occupational similarity channel, *OCC*, a one standard deviation increase in the presence of occupationally-similar local industries (25.47) leads to a 14.8% increase in city industry growth when using the results from Column 6 of Table 2.2. Thus, both of these channels appear to exert a substantial positive effect on city-industry growth.

Table 2.2: Results with all cross-industry spillover channels

	(1)	(2)	(3)	(4)	(5)	(6)
IOin1	0.0170 (0.0114)	0.0181 (0.0122)	0.0156 (0.0114)	0.0645*** (0.0187)	0.0565*** (0.0191)	0.0587*** (0.0190)
IOout1	0.0055 (0.0049)	0.0003 (0.0052)	-0.0010 (0.0049)	-0.0079 (0.0113)	-0.0145 (0.0114)	-0.0187 (0.0119)
EMP	0.0001 (0.0005)	0.0001 (0.0006)	-0.0002 (0.0006)	0.0010 (0.0016)	0.0027* (0.0014)	0.0016 (0.0014)
OCC2	0.0016 (0.0014)	0.0012 (0.0013)	0.0010 (0.0013)	0.0061** (0.0030)	0.0066** (0.0033)	0.0058* (0.0033)
within	-0.0437*** (0.0161)	-0.0258* (0.0140)	-0.0246* (0.0143)			
Observations	4,253	3,544	3,539	4,253	3,539	3,539
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
Within terms	homog	homog	homog	heter	heter	heter
KP under		24.86	25.45		22.09	24.52
KP weak		4677.9	858.6		52.36	35.68

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heterogeneous regressors *within* are included in Columns 4-6 but not displayed. City-time and industry-time effects are included in all regressions but not displayed. 2SLS regressions use lagged instruments. Note that the number of observations falls for the instrumented regressions because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*. "KP under id." denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). "KP weak id." denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic.

Our analysis can also help us understand the strength of within-industry spillovers,

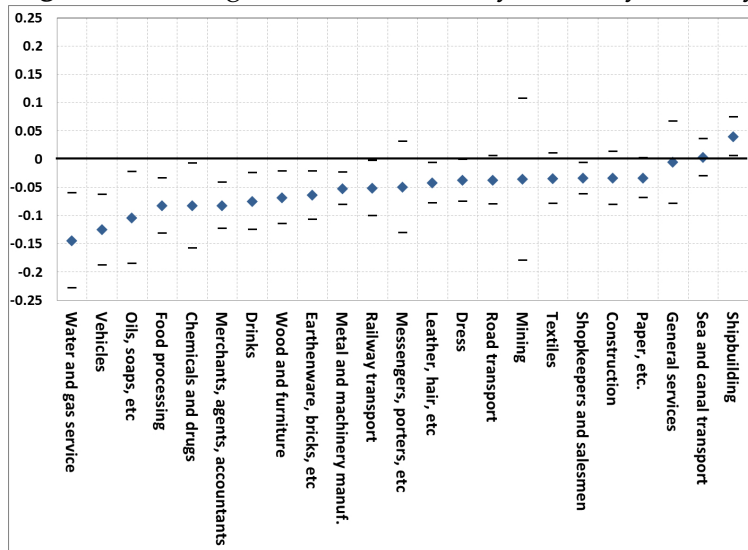
reflected in the $\ln(L_{ict})$ term in Equation 2.17.⁴¹ When analyzing these results, it is important to keep in mind that they reflect the *net* effect of within-industry agglomeration forces, which may be generated through a balance between agglomeration forces and negative forces such as competition or mean-reversion due to the diffusion of technologies across cities. We cannot identify the strength of local within-industry agglomeration forces independent of counteracting forces. However, it is the net strength of these forces, which we are able to estimate, that is relevant for understanding the contribution of within-industry agglomeration forces to city growth. Thus, our results suggest that within-industry agglomeration effects generally do not make a positive contribution to city employment growth.

We have already seen, in Table 2.2 Columns 1-3, that the average within-industry effect across all industries is negative. These results are consistent with negative dynamic within-industry effects, perhaps linked to the unwillingness of firms to share new ideas with their direct competitors. However, the fact that our results change substantially once we allow for heterogeneous within-industry effects, as in Columns 4-6 of Table 2.2, suggests that these are likely to vary substantially across industries. We explore these heterogeneous within-industry effects in Figure 2.2, which presents coefficients and 95% confidence intervals for industry-specific within-industry spillover coefficients from regressions corresponding to Column 6 of Table 2.2.

In only one industry, shipbuilding, do we observe any evidence of positive within-industry effects. This industry was characterized by increasing returns and strong patterns of geographic concentration. All other industries exhibit slower growth in locations where initial industry employment was large, after controlling for other forces. Within-industry agglomeration benefits, it would appear, are more the exception than the rule.

⁴¹In a static context these are often referred to as localization economies.

Figure 2.2: Strength of within-industry effects by industry



Results correspond to the regression described in Column 6 of Table 2.2. This figure displays coefficient estimates and 95% confidence intervals based on standard errors clustered by city-industry, city-year, and industry-year. The regression includes a full set of city-year and industry-year effects as well as *between* terms. Both the within and between terms are instrumented using one-decade lags.

The results presented so far describe coefficients generated using all industries, where each industry is given equal weight. We have also calculated weighted regressions, where the set of observations for each city-industry is weighted based on employment in that city-industry at the beginning of each period. These results, available in Appendix 2.3, show qualitatively similar results to those shown above for the IOin term, with only slightly smaller estimated coefficients. This provides confidence that our main findings are not being driven by small cities or industries. The weighted results also show stronger evidence of a negative effect on the IOout term, but this finding appears to be quite sensitive to the set of industries included in the analysis. The agglomeration benefits from occupationally similar industries disappear when weighting by city-industry size, suggesting that labor market pooling benefits may be larger for small industries or in small cities.

We have also investigated the robustness of our results to dropping individual industries or individual cities from the analysis database (see Appendix 2.3). These exercises

show that the significance of the estimates on the *IOin* and *OCC* channels are robust to dropping any city or any industry. However, the estimated coefficient and confidence levels for the *IOout* coefficient is sensitive to the exclusion of particular industries. Specifically, when shipbuilding is excluded we observe that the *IOout* coefficient becomes positive but not statistically significant.⁴² This suggests that in general the presence of local buyers may have a mild positive effect on industry growth.

In addition, we have explored the sensitivity of our results to using alternative functional forms to represent the relationship between spillovers and technological progress. In Appendix 2.3 we present results using alternative concave functional relationships such as a square root or fifth root. Our findings are not sensitive to these alternatives.

We have also explored the robustness of our results to the use of alternative connections matrices. In particular, in Appendix 2.3 we present results obtained while using the less detailed input-output table constructed by Horrell *et al.* (1994), which covers 12 more aggregated industry categories in 1841. When using this alternative matrix we continue to find evidence of positive effects generated by the presence of local suppliers. These results also suggest that local buyers may generate positive benefits, but as before this result appears to be sensitive to the set of industries included in the analysis.

The results discussed so far reveal average patterns across all industries. An additional advantage of our empirical approach is that it is also possible to estimate industry-specific coefficients in order to look for (1) heterogeneity in the industries that benefit from each type of inter-industry connection or (2) heterogeneity in the industries that produce each type of inter-industry connections. In Appendix 2.3, we estimate industry-specific coefficients for both spillover-benefiting and spillover-producing industries and then compare them to a set of available industry characteristics such as firm size, export and final goods sales shares, and labor or intermediate cost shares. With only 23 estimated industry coefficients we cannot draw strong conclusions from these relationships. However, our results

⁴²Shipbuilding stands out relative to the other industries because it is particularly reliant on local geography.

do suggest several interesting patterns. The only clear result is that industries that benefit from or produce spillovers through the OCC channel tend to have a higher labor cost to sales ratio, a finding that seems very reasonable. We also observe a consistent negative relationship between firm size and all types of inter-industry connections. While this relationship is not statistically significant, it is consistent across all spillover types and it fits well with previous work highlighting the importance of inter-industry connections for smaller firms (e.g., [Chinitz \(1961\)](#)).

We can also look at how the estimated industry-specific within-industry coefficients are related to industry characteristics. This is done in [Appendix 2.3](#). With such a small number of industry coefficients we cannot draw strong conclusions from these results. However, we do observe some evidence that within-industry connections are more important in industries with larger firm sizes, which contrasts with the consistent negative relationship that we observe between firm size and cross-industry spillovers.

While the analysis described above focuses on spillovers occurring within-cities, we have also explored the possibility that there may be important cross-city effects. To explore cross-city effects, we have run additional regressions including variables measuring market size as well as cross-industry spillovers occurring across cities. Our results, reported in [Appendix 2.3](#), suggest that cross-city effects are much weaker than within-city forces. This makes sense given that we think that the shape of cities reflects the rapidly decaying strength of local agglomeration forces. We also find that accounting for cross-city effects has little impact on our estimates of the strength of within-city agglomeration forces.

2.6 STRENGTH OF THE AGGLOMERATION FORCES

In this section we examine the relationship between city size and city-industry growth and show how our city-year effects can be used to construct a summary measure of the

aggregate strength of the many cross-industry agglomeration forces present in our model. In standard urban models, the impact of agglomeration forces is balanced by congestion forces related to city size, operating through channels such as higher housing prices or greater commute times. In our model, we have been largely agnostic about the form of the congestion forces, which will be captured primarily by the city-time effects. Thus, examining these estimated city-time coefficients offers an opportunity for assessing the *net* impact of dynamic congestion or agglomeration force related to overall city size.⁴³ Also, the difference between these estimated city-time effects and city growth rates must be due to the impact of the agglomeration forces in the estimation equation. As a result, comparing the estimated city-time effects to actual city growth rates allows us to quantify the combined strength of the many cross-industry agglomeration forces captured by our measures.

To gain some intuition into this comparative exercise, consider the graphs in Figure 2.3. The dark blue diamond symbols in each graph describe, for each decade starting in 1861, the relationship between the actual growth rate of city working population and the log of city population at the beginning of the decade. The slopes of the fitted lines for these series fluctuate close to zero, suggesting that on average Gibrat's Law holds for the cities in our data.

We want to compare the relationship between city size and city growth in the actual data, as shown by the dark blue diamonds in Figure 2.3, to the relationship between these variables obtained while controlling for within and cross-industry agglomeration forces. This can be done using the estimated city-time effects represented by θ_{ct} in Eq. 2.18. The red squares in Figure 2.3 describe the relationship between the estimated city-year coefficients for each decade, $\hat{\theta}_{ct}$, and the log of city population at the beginning of each decade. In essence, these are showing us the relationship between city size and city growth after controlling for national industry growth trends and the agglomeration forces

⁴³These results will reflect only the net impact of city size, including both congestion and agglomeration forces.

included in our model. We can draw three lessons from these graphs. First, in all years the fitted lines based on the $\hat{\theta}_{ct}$ terms slope downward more steeply than the fitted lines for actual city growth. This suggests that, once we control for cross-industry agglomeration forces, city size is negatively related to city growth, consistent with the idea that there are dynamic city-size congestion forces. Second, the difference between the slopes of the two fitted lines can be interpreted as the aggregate effect of the various agglomeration forces in our model averaged across cities. Put simply, if we can add up the strength of the convergence force in any period and compare it to the actual pattern of city growth, then the difference must be equal to the strength of the agglomeration forces. Third, the patterns described in Figure 2.3 appear to be close to linear in logs, suggesting that these forces do not differ dramatically across different city sizes.

The strength of these effects can be quantified in terms of the implied convergence rate following the approach of Barro & Sala-i Martin (1992). We run,

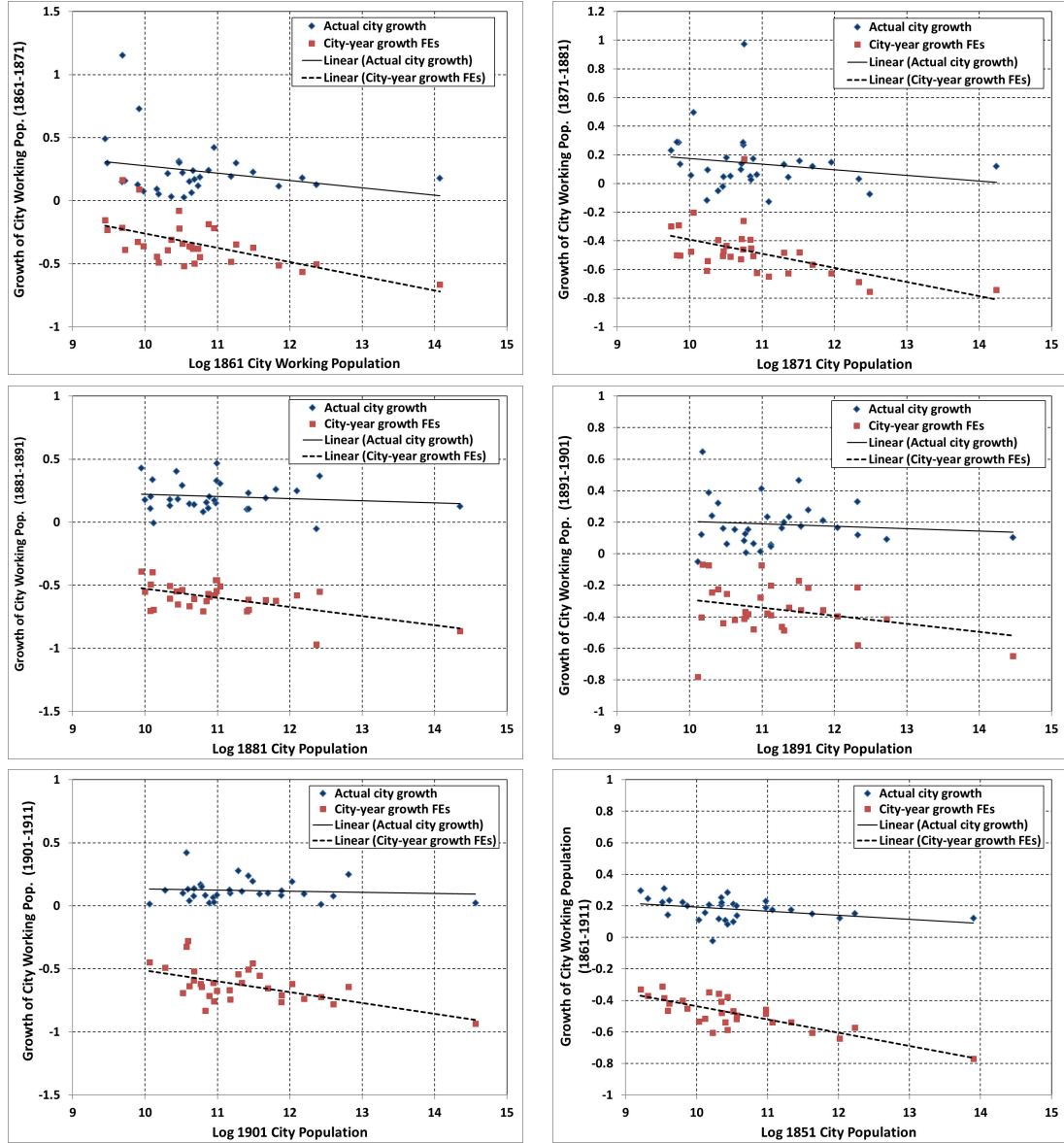
$$\hat{\theta}_{ct} = a_0 + a_1 \ln(L_{ct}) + \epsilon_{ct} \quad (2.19)$$

$$\tilde{\theta}_{ct} = b_0 + b_1 \ln(L_{ct}) + \epsilon_{ct} \quad (2.20)$$

where $\hat{\theta}_{ct}$ is the estimated city-time effect for the decade from t to $t + 1$ from a regression based on Eq. 2.18 (but omitting the within terms, which clearly represent a convergence rather than a divergence force), L_{ct} is the working population of the city in year t , and $\tilde{\theta}_{ct}$ is the industry-demeaned growth rate of city c from t to $t + 1$.⁴⁴ These regressions are run separately for each decade from 1861 to 1911, either with or without weighting each observation by initial city-industry employment, and using lagged values as instruments as in the main results. Convergence rates are then calculated using the estimated a_1 and b_1 coefficients. A comparison of the a_1 and b_1 coefficients describes, at the city level, the

⁴⁴I.e., $\tilde{\theta}_{ct}$ is the estimated value of θ_{ct} obtained from the regression $\Delta \ln(L_{ict+1}) = \theta_{ct} + \chi_{it} + e_{ict}$.

Figure 2.3: City size and city growth



Solid lines: Fitted lines comparing actual city growth over a decade to the log of city size at the beginning of the decade. **Dotted lines:** Fitted lines comparing estimated coefficients from city-time effects for each decade to the log of city size at the beginning of the decade. **Blue diamonds:** Plot the actual city growth over a decade against the log of city population at the beginning of the decade. **Red squares:** Plot the estimated city-time coefficients over the same decade (the θ_{ct} terms estimated using Eq. 2.18) against the log of city population at the beginning of the decade. The bottom right-hand panel compares the log of city population in 1851 to the average of city growth rates over the entire 1861-1911 period and the average of city-time fixed effects across the entire 1861-1911 period.

impact of accounting for cross-industry spillovers.

Results based on unweighted regressions are presented in the top panel of Table 2.3. The two left-hand columns describe the results from Equation 2.19 and the annualized city-size divergence rate implied by these estimates. The next two columns describe similar results based on Equation 2.20. The difference between these two city-size divergence rates, given in the right-hand column, describes the aggregate strength of the agglomeration force reflected in the cross-industry terms. These results suggest that the strength of city agglomeration forces, in terms of the implied divergence rate, was 2.0-2.3% per decade. In the bottom panel of Table 2.3 we calculate similar results except that the $\hat{\theta}_{ct}$ terms are obtained using regressions in which each observation is weighted based on the employment in the city-industry at the beginning of each period. These results suggest a slightly weaker agglomeration force, equal to an implied divergence rate of 1.6-1.7% per decade.

Table 2.3: Measuring the aggregate strength of the agglomeration forces

Results based on un-weighted regressions					
	Results based on θ_{ct}		Results for actual city growth		Aggregate strength of agglomeration force (implied divergence rate per decade)
	Estimated city-size coefficient	Implied divergence Beta	Estimated city-size coefficient	Implied divergence Beta	
1861-1871	-0.076	7.86%	-0.056	5.71%	2.15%
1871-1881	-0.062	6.38%	-0.042	4.26%	2.12%
1881-1891	-0.035	3.60%	-0.015	1.53%	2.07%
1891-1901	-0.014	1.43%	0.006	-0.61%	2.04%
1901-1911	-0.046	4.70%	-0.023	2.37%	2.33%

Results based on regressions weighted by city-industry size in 1851					
	Results based on θ_{ct}		Results for actual city growth		Aggregate strength of agglomeration force (implied divergence rate per decade)
	Estimated city-size coefficient	Implied divergence Beta	Estimated city-size coefficient	Implied divergence Beta	
1861-1871	-0.066	6.86%	-0.051	5.23%	1.63%
1871-1881	-0.052	5.29%	-0.036	3.61%	1.67%
1881-1891	-0.037	3.79%	-0.021	2.13%	1.66%
1891-1901	-0.026	2.64%	-0.010	1.00%	1.64%
1901-1911	-0.018	1.86%	-0.002	0.20%	1.65%

Column 1 presents the a_1 coefficients from estimating Equation 2.19 for each decade (cross-sectional regressions). Column 2 presents the decadal convergence rates implied by these coefficients. Column 3 presents the b_1 coefficients from estimating Equation 2.20 and Column 4 presents the decadal divergence rates implied by these coefficients. Column 5 gives the aggregate strength of the divergence force due to the agglomeration economies, which is equal to the difference between the decadal divergence coefficients in Columns 2 and 4. Results in the top panel are unweighted, while results in the bottom panel are from regressions in which each city-industry observation is weighted by the employment in that city-industry at the beginning of the period.

We can use a similar exercise to estimate the aggregate strength of the convergence force due to within-industry effects. We begin by estimating,

$$\Delta \ln(L_{ict+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \theta_{ct}^{WITHIN} + \chi_{it} + e_{ict}. \quad (2.21)$$

which is just Eq. 2.18 with the cross-industry terms omitted. Next, we use the estimated values of θ_{ct}^{WITHIN} to estimate,

$$\hat{\theta}_{ct}^{WITHIN} = d_0 + d_1 \ln(L_{ct}) + \epsilon_{ct}. \quad (2.22)$$

We then calculate the convergence force associated with the within-industry terms using the same approach that we used previously, i.e. we compare the d_1 coefficients with the slopes estimated using Eq. 2.20. Table 2.4 describes the results. The negative measured divergence force in this table highlights that within-industry effects, on net, act as a convergence force. The strength of this force is sensitive to whether the regressions are weighted, which suggests that the negative within-industry employment effects are likely to vary with initial city-industry employment.

Table 2.4: Measuring the aggregate strength of the convergence force associated with the within-industry effects

Results based on un-weighted regressions					
	Results based on θ^{WITHIN}		Results for actual city growth		Aggregate strength of agglomeration force (implied divergence rate per decade)
	Estimated city-size coefficient	Implied divergence Beta	Estimated city-size coefficient	Implied divergence Beta	
1861-1871	-0.004	0.43%	-0.056	5.71%	-5.28%
1871-1881	0.009	-0.87%	-0.042	4.26%	-5.13%
1881-1891	0.034	-3.38%	-0.015	1.53%	-4.91%
1891-1901	0.056	-5.41%	0.006	-0.61%	-4.80%
1901-1911	0.025	-2.48%	-0.023	2.37%	-4.85%

Results based on regressions weighted by city-industry size in 1851					
	Results based on θ^{WITHIN}		Results for actual city growth		Aggregate strength of agglomeration force (implied divergence rate per decade)
	Estimated city-size coefficient	Implied divergence Beta	Estimated city-size coefficient	Implied divergence Beta	
1861-1871	-0.048	4.91%	-0.051	5.23%	-0.33%
1871-1881	-0.033	3.38%	-0.036	3.61%	-0.24%
1881-1891	-0.019	1.92%	-0.021	2.13%	-0.21%
1891-1901	-0.008	0.81%	-0.010	1.00%	-0.20%
1901-1911	0.000	-0.01%	-0.002	0.20%	-0.21%

Column 1 presents the d_1 coefficients from estimating Equation 2.22 for each decade (cross-sectional regressions). Column 2 presents the decadal divergence rates implied by these coefficients. Column 3 presents the b_1 coefficients from estimating Equation 2.20 and Column 4 presents the decadal divergence rates implied by these coefficients. Column 5 gives the aggregate strength of the divergence force due to the agglomeration economies, which is equal to the difference between the decadal convergence coefficients. The negative values in Column 5 indicate that within-industry effects are, on net, a source of convergence across cities. Results in the top panel are unweighted, while results in the bottom panel are from regressions in which each city-industry observation is weighted by the employment in that city-industry at the beginning of the period.

There are some caveats to keep in mind when assessing these results. First, there are likely to be agglomeration forces not captured by our estimation. These omitted agglom-

eration forces may be partially reflected in the city-year fixed effects, which would lead us to understate the strength of the agglomeration forces. Second, some congestion forces may also be captured by our cross-industry terms. Similarly, there may be some agglomeration forces captured by the within-industry terms, which will also not be reflected in our results. Thus, the strength of the cross-industry agglomeration force measured here is likely to be a lower bound on the true values.

2.7 CONCLUSION

In the introduction, we posed a number of questions about the nature of localized agglomeration forces. The main contribution of this study is to provide a theoretically grounded empirical approach that can be used to address these questions and the detailed city-industry panel data needed to implement it. We can now provide some answers for the particular empirical setting that we study. First, we find evidence that cross-industry agglomeration economies were more important than within-industry agglomeration forces for generating city employment growth. Within-industry effects are, on net, generally negative. This suggests that local clusters of firms working in the same industry, which have attracted substantial attention, are unlikely to deliver dynamic benefits. Second, our results suggest that industries grow more rapidly when they co-locate with their suppliers or with other industries that use occupationally-similar workforces. This result is in line with arguments made by [Jacobs \(1969\)](#), as well as recent empirical findings. We document a clear negative relationship between city size and city growth that appears once we account for agglomeration forces related to a city's industrial composition. This suggests that Gibrat's law is generated by a balance between agglomeration and dispersion forces. An estimate of the overall strength of the agglomeration forces captured by our approach, in terms of the implied annual divergence rate in city size, is 1.6-2.3% per decade.

The techniques introduced in this paper can be applied in any setting where sufficiently rich long-run city-industry panel data can be constructed. Recent work has made progress in constructing data of this type for the U.S. in both the modern and historical period. Applying our approach to these emerging data sets is another promising avenue for future work.

Chapter 3

Gravity estimation with unobserved bilateral flow data

3.1 INTRODUCTION

How to estimate unobserved bilateral trade flows? Understanding trade patterns between countries has traditionally been the main interest of the International Trade literature. Accordingly, one of the key data requirements to estimate trade models is precisely bilateral trade data. However, over time the boundaries between the fields of International Trade and Regional Economics are increasingly blurred. Models that were originally developed to explain international trade patterns are now commonly adapted to study economic questions at the subnational level. Yet, the domestic counterpart of international trade data (i.e. data on trade flows between subnational units) is typically not available. This paper proposes a novel methodology to approximate bilateral trade flows by sector that is consistent with a large class of standard gravity models.

At the national level we now have long and finely disaggregated time series on international trade flows. The bilateral nature of these trades implies that a complete matrix of flows can be constructed by drawing data from either side of the trade, or both. Therefore, the quality of such data is fairly consistent across developed as well as developing countries. However, at the subnational level there is a lot more heterogeneity in standards and data availability. The US is a rare exception in that it collects and publishes detailed domestic trade data, but it only began doing so in recent years. Perhaps surprisingly, many other developed countries that normally have high-quality data in other domains, do not collect or publish any domestic trade data (e.g. France, to the best of my knowledge).¹ The lack of detailed domestic trade data is even more pronounced in developing countries. This precludes many interesting empirical applications of regional trade models and it is probably one of the reasons why the Regional Economics literature has focused primarily on developed countries.

¹Tax authorities in countries with value-added taxes have the raw data that, if aggregated, could be used to compute domestic trade flows with very high spatial and sectoral resolution. Unfortunately such data is rarely made available to researchers. See [Dhyne et al. \(2015\)](#) for an example from Belgium. I thank Jonathan Dingel for pointing out this potential data source.

In this paper I adapt the methodology by [Miscio & Soares \(2016\)](#) to predict sectoral domestic flows between Brazilian metropolitan areas. This methodology was initially developed to infer *commuting* flows from aggregate data on population by place of residence and by place of work. I show that it can also be applied to infer sectoral domestic trade flows and I use it to compute the domestic trade flows used in chapter 1. A related approach to infer commuting and domestic trade flows can be found respectively in [Ahlfeldt *et al.* \(2015\)](#) and [Monte *et al.* \(2015\)](#). Their methodologies differs from the one presented here in that they require stronger assumptions and deliver weaker results. Therefore, the approach proposed by [Miscio & Soares](#) is more general and more widely applicable than the alternatives currently available.

In section 3.2 I present [Miscio & Soares](#)'s methodology adapted to a trade context and how it compares with alternative approaches. In section 3.3 I test its predictive power using US data from the 2012 Commodity Flow Survey. Finally, I apply it to Brazilian data and discuss its limitations. Section 3.4 concludes.

3.2 METHODOLOGY

In this section I show how to estimate the parameters of a general gravity equation without using bilateral flow data.² I start by laying out the equations of a Poisson estimator applied to a general gravity equation. Then I show how to interpret these equations as moment conditions of a General Method of Moments estimator. Finally I discuss how to modify these moments in order to relax the data requirements. I conclude the section with a discussion of alternative approaches found in the literature.

As a starting point, consider the large class of trade models that yield what [Head & Mayer \(2014\)](#) call a *general gravity equation*. This includes, among others, [Armington \(1969\)](#), [Eaton & Kortum \(2002\)](#) and [Melitz \(2003\)](#) with the assumption of a Pareto dis-

²This section draws heavily on [Miscio & Soares \(2016\)](#), whose manuscript is not circulating yet.

tribution of firm productivities, as shown in [Chaney \(2008\)](#). In all these trade models bilateral flows can be expressed generically as

$$x_{od} = \alpha_o \beta_d f(\tau_{od}) \quad (3.1)$$

where x_{od} is the demand in region d (d for destination) for goods produced in region o (o for origin), α_o is any origin-specific effect, β_d is any destination-specific effect. For example in the Armington model α_o is a function of the labor cost in o and β_d is a function of total income and the price index in d . The term $f(\tau_{od})$ is an origin-destination specific trade friction. In empirical applications $f(\tau_{od})$ is usually parameterised as a function of distance between o and d or as a dummy for whether the origin-destination pair belongs to a regional trade block, has a common language, etc. Since the focus in this paper is on domestic trade flows, I restrict my attention to distance only. Therefore, as common in the trade literature, I assume that $f(\tau_{od}) = \tau_{od}^\gamma$, where τ denotes distance.

Gravity equations like 3.1 are typically estimated as linear regressions after taking logs on both sides. However, [Silva & Tenreyro \(2006\)](#) show that this practice leads to a number of problems in the presence of heteroskedasticity or if the full matrix of bilateral flows includes a large number of zeros, as is typically does. They advocate instead the use of a Poisson estimator, which can account for both issue and is naturally suited for count data. Other authors have documented desirable properties of Poisson estimators when used to estimate gravity equations. For instance, [Arvis & Shepherd \(2013\)](#) and [Fally \(2015\)](#) show that Poisson estimators with importer fixed effects and exporter fixed effects satisfy *adding-up* constraints.³ On the other hand, log-linear estimators fail this test. Accordingly, we take a Poisson estimator as the building block of our methodology.

For illustration purposes only, in the remainder of this section I will use a naive gravity equation (as defined in [Head & Mayer \(2014\)](#)) where I replace the full set of fixed

³An example of an adding-up constraint is the requirement that the sum of all estimated export flows from a given origin be equal to the sum of all observed export flows from that origin.

effects with origin and destination GDP and a constant, respectively y_o, y_d, κ . I re-write the gravity equation 3.1 as

$$x_{od} = \kappa y_o^\alpha y_d^\beta \tau_{od}^\gamma = e^{\log \kappa + \alpha \log y_o + \beta \log y_d + \gamma \log \tau_{od}} \equiv e^{(\log w)'\theta} \quad (3.2)$$

where $\log w$ is the vector of regressors y_o, y_d, τ_{od} and θ is the set of parameters $\kappa, \alpha, \beta, \gamma$. Poisson estimators are derived from the assumption that the conditional mean is correctly specified. That is,

$$\mathbb{E}(x_{od}|w) = e^{(\log w)'\theta} \quad (3.3)$$

The first order conditions (FOC) from likelihood maximization are

$$\sum_o \sum_d \left[\left(x_{od} - e^{(\log w)'\theta} \right) \cdot (\log w) \right] = 0 \quad (3.4)$$

Notice that these FOC can also be interpreted as moment conditions of a General Method of Moment (GMM) estimator, where the identifying assumption is that the prediction error (i.e. the difference between actual and predicted flows) is mean independent from the instruments contained in the vector $\log w$.⁴ We will come back to this interpretation later. Specifically, and after rearranging the terms, the FOC are:

$$\left\{ \begin{array}{ll} \sum_o \sum_d x_{od} \cdot 1 = \sum_o \sum_d e^{(\log w)'\theta} \cdot 1 & \text{FOC identifying } \kappa \\ \sum_o (\log y_o \cdot \sum_d x_{od}) = \sum_o \sum_d e^{(\log w)'\theta} \cdot \log y_o & \text{FOC identifying } \alpha \\ \sum_d (\log y_d \cdot \sum_o x_{od}) = \sum_o \sum_d e^{(\log w)'\theta} \cdot \log y_d & \text{FOC identifying } \beta \\ \sum_o \sum_d x_{od} \cdot \log \tau_{od} = \sum_o \sum_d e^{(\log w)'\theta} \cdot \log \tau_{od} & \text{FOC identifying } \gamma \end{array} \right.$$

⁴Strictly speaking, the moment conditions would also feature the term $\frac{1}{OD}$ multiplying the left-hand side, where O and D are the total number of origins and destinations. However, in empirical applications this is irrelevant because a researcher always works with finite data and numerical solutions are computed up to a finite number of decimal digits. For example, suppose we set the tolerance to 10^{-6} and that we have a total of 10^3 observations. Multiplying the left-hand side by $\frac{1}{10^3}$ is the same as requiring that the level of tolerance is 3 orders of magnitude bigger.

Each equation allows us to identify one of the parameters conditional on data and on a value for each one of the other parameters. Jointly, the system of FOC identifies all the parameters. Notice that in the first three equations we do not require data on all the bilateral flows x_{od} , but simply their partial sums $\sum_o \sum_d x_{od}$, $\sum_d x_{od}$, $\sum_o x_{od}$, i.e. respectively the sum across all origins and destinations, the sums across all destinations and the sums across all origins. This observation drastically simplifies the data requirements of the estimator. In the international trade context, the first of these three partial sums is global output (or, equivalently, global demand), the second one is total production in country o and the third one is total demand in country d . Notice also that none of these equations implies that trade needs to be balanced on a country by country basis. This is particularly useful when working with data at the subnational level because, even more than in an international context, domestic trade may be systematically unbalanced due, for instance, to federal redistribution policies. The first FOC also shows how Poisson estimators satisfy adding-up constraints. In this case, κ is estimated such that the sum of all observed flows equals the sum of all estimated flows.

Nonetheless, unlike the first three FOC the last equation requires actual bilateral flow data, not just partial sums. Therefore, in the absence of such data we cannot implement a Poisson estimator. However, we can return to the interpretation of the FOC as moment conditions. One advantage of the GMM interpretation is that it allows us to complete the system of equations by replacing the last one with alternative moments that identify the same parameter.⁵ The challenge is to find moments in the data that identify γ but do not require the entire matrix of flows x_{od} . [Miscio & Soares \(2016\)](#) propose several alternatives that differ in their data requirements and the underlying identifying assumption.⁶ In the simplest of such alternatives we calibrate γ to a plausible value. Then the partial

⁵On the other hand, we cannot replace one of the FOC and still call it a Poisson estimator as that would imply a different likelihood function.

⁶These alternative moment conditions cover cases where the researcher only observes internal trade flows, or bilateral flows at a higher level of spatial aggregation. Finally, one of the alternatives is a set of moment inequalities that relies on the observation that on average internal trade is larger than bilateral trade.

sums of trade flows are sufficient to identify all the other parameters using the remaining equations. The *adding-up* property is still satisfied even after we replace the last equation with one of the alternative moment conditions.

The discussion above extends to the general case where instead of variables y_o, y_d and a constant we have a full set of origin and destination fixed effects. That is, with N regions we need $2N$ equations to identify $2N$ parameters. Of these, $2N - 1$ equations identify the fixed effects and do not require bilateral flow data. Only the equation identifying the distance elasticity requires bilateral flow data and we replace it with one of the alternative moment conditions as discussed above. This also extends to sectoral gravity equations, such as the one found in [Caliendo & Parro \(2015\)](#) and [Caliendo *et al.* \(2015\)](#). Again, the fact that we do not need to impose any trade balance condition is a clear advantage as the very nature of comparative advantage suggests that on a sector by sector basis trade flows are unbalanced.

The closest alternative methodology to estimate gravity parameters without bilateral flow data is in a paper by [Ahlfeldt *et al.* \(2015\)](#). They develop a spatial equilibrium model based on [Eaton & Kortum \(2002\)](#) and obtain a gravity equation describing commuting flows between city districts. Their gravity equation also belongs to the class of *general gravity* discussed above. [Ahlfeldt *et al.*](#) 's goal is to estimate all the gravity parameters using data from Berlin in three different years: pre-World War II, Berlin-wall period, post Berlin-wall-period. However, they only have commuting data for the last period. Therefore, they use data from the post Berlin-wall period and a linear fixed effects estimator to estimate the distance elasticity. Then, for all years they set the distance elasticity parameter at the estimated value and use equations from the model and other data such as district population and wages to estimate the remaining parameters. A similar estimation strategy is found in [Monte *et al.* \(2015\)](#). They develop a different spatial equilibrium model, also based on [Eaton & Kortum \(2002\)](#) and obtain a gravity equation for trade flows that also belongs to the class of *general gravity*. Their goal is to estimate the gravity pa-

rameters at the level of US Counties but they only observe trade data at a higher level of spatial aggregation, i.e. CFS regions. Their strategy, as in the previous paper, is to use the available bilateral trade data to estimate a distance elasticity using a fixed effects estimator. Then, they set the distance elasticity parameter to the estimated value and use county-level data along with equilibrium conditions from the model in order to estimate the remaining gravity parameters at the county level.

The estimation strategies in [Ahlfeldt *et al.* \(2015\)](#) and [Monte *et al.* \(2015\)](#) compare to the one presented in this paper in the following way. First, their estimating equations are model-dependent, i.e. they are derived from a fully specified general equilibrium model that, among other things, delivers a gravity equation. The estimating equations in [Miscio & Soares \(2016\)](#) are consistent with *any* micro-foundation that delivers a general gravity equation. Therefore, our approach is more general as it requires less assumptions. Second, their solution to estimate the distance elasticity parameter using observed bilateral flows and then holding the parameter at that value in subsequent estimations is equivalent to the simplest of our approaches to complete the system of equations, i.e. parameterizing γ to a reasonable value.

[Ahlfeldt *et al.*](#) and [Monte *et al.*](#) 's lack of generality can be illustrated with the following example. Suppose that a researcher requires predicted trade flows by sector instead of aggregate trade flows, or commuting flows by worker type rather than aggregate commuting flows, then neither the strategy in [Ahlfeldt *et al.* \(2015\)](#) nor the one in [Monte *et al.* \(2015\)](#) can be used. Both strategies depend on models that only describe aggregate flows. On the other hand, our strategy can still be used. Indeed, in the next section I apply it to estimate sectoral domestic flows using US and Brazilian data.

3.3 APPLICATIONS

In this section I use the methodology presented above to estimate sectoral domestic trade using US and Brazilian data. The exercise on US data is meant to be a way to validate the methodology. In particular, I show how the predicted sectoral flows between CFS regions compare to observed flows. Then, I apply the methodology to Brazilian data in order to predict flows between regions analogous to the US CFS regions and for which there is no trade data.

3.3.1 Estimation of US sectoral domestic trade flows

In the US, the most recent Commodity Flow Surveys provides detailed information on internal shipments of goods by sector at the State level, as well as at the CFS-region level. CFS regions are a particular aggregation of Counties that can be thought of as the union of metropolitan areas and State remainders.⁷ Earlier surveys provided less sectoral or geographical detail. For instance the earliest Commodity Flow Survey from 1993 only provided total flows between States (i.e. not disaggregated by industry). On the other hand, sectoral flows between US metropolitan areas are in the public domain only starting from 2007.⁸

The availability of both levels of spatial aggregation is particularly useful in that it allows me to test the exact methodology that I use in chapter 1 and in the next section to infer sectoral trade flows between Brazilian metropolitan areas. In the public CFS data a number of observations are censored for confidentiality reasons. Since the degree of

⁷CFS areas correspond to State components of Consolidated Statistical Areas defined by the Office of Management and Budget, and selected Core-based Statistical Areas (e.g. the New York-Newark CSA stretches four States and it is therefore divided into four components). All remaining Counties are merged into State residuals.

⁸Additional geographic and sectoral detail is available in the confidential CFS micro-data from 1997 onward. In particular, the micro-data contains establishments' location down to the zipcode. However, the sampling frame used in the survey and the resulting sampling weights do not allow researchers to infer bilateral flows below the CFS-area level. In this section by *metropolitan areas* I mean the particular aggregation of Counties used in CFS publications.

censoring increases with spatial and sectoral detail, I only focus on highly aggregated industries, i.e. mining, manufacturing and wholesale trade, as well as selected manufacturing subsectors and the total across all sectors.⁹

First, I estimate a standard Poisson gravity equation analogous to equation 3.1 with a full set of origin and destination fixed effects. I perform a separate estimation for each sector and for each of the two levels of spatial aggregation (States and CFS-areas). Bilateral distances are calculated as great-circle distances between population-weighted centroids. Internal distances are computed as in Head & Mayer (2000).¹⁰ Figure 3.1 shows the estimated distance parameters.

The first observation is that the estimated distance elasticities differ across sectors. However, across all sectors the choice of the spatial unit does not seem to affect the point estimates as they all fall close to the 45 degree line. I use the estimated parameters from the Poisson regressions at the CFS-area level and equation 3.1 to compute predicted bilateral trade flows by sector. The correlations between actual and predicted flows from these regressions are the benchmark against which I compare the predictive power of the methodology proposed in this paper. These benchmark correlations are displayed in figure 3.2 (the other values shown in the same figure are explained later in this section).

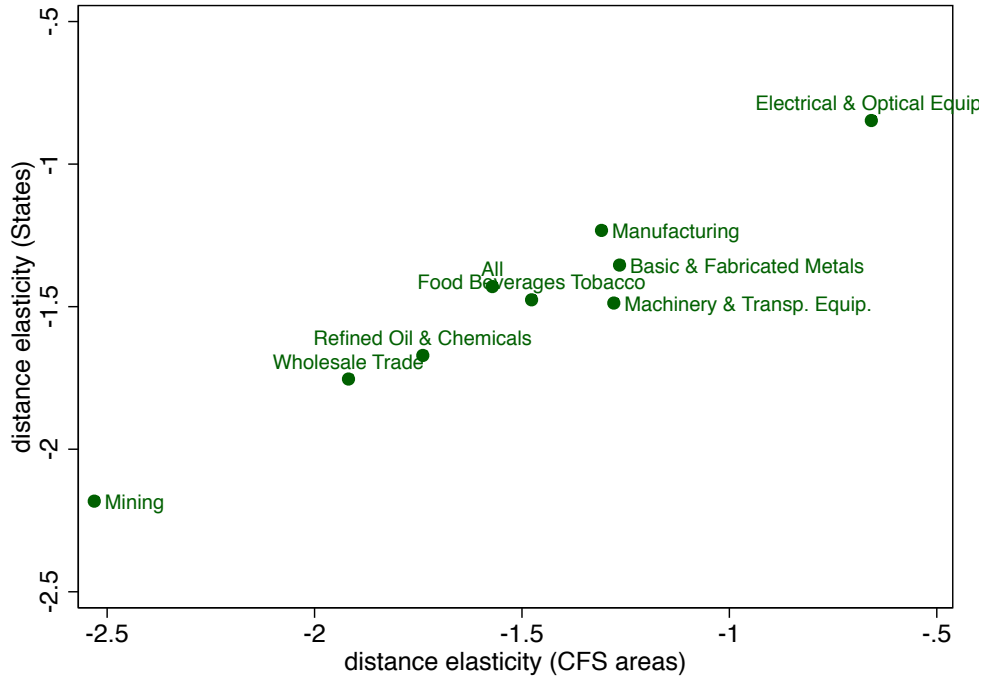
The correlations based on the Poisson estimates are taken as the benchmark because they are the most data intensive and the best approximation one can get to actual flows using any trade model that belongs to the class of general gravity as defined earlier.¹¹

⁹For instance at the higher levels of sectoral aggregation (mining, manufacturing, wholesale trade and total trade across all sectors) no observation is censored at the CFS-area level. On the other hand, in each of the manufacturing subsectors the rate of censoring is double-digit.

¹⁰Specifically, for each spatial unit, whether States or CFS areas, I calculate the average longitude and latitude as the weighted average of Counties' centroids using Counties' population as weights. These population-weighted coordinates are then used to compute great-circle bilateral distances. Internal distances are computed as the average radial distance of points in a circle with the same area as the spatial units, i.e. $\tau_{ii} = \frac{2}{3} \sqrt{\frac{A_i}{\pi}}$, where A_i is, for instance, the land area of a CFS region. I also used the average shipment mileage provided with the CFS data and the results are comparable. I prefer the former approach as it is the exact method I use on Brazilian data, where I do not have average shipment mileage.

¹¹By *best approximation* I mean two things: (i) Poisson estimators are the only maximum-likelihood estimators that satisfy adding up constraints, as shown in Fally (2015), and therefore any partial sum of actual flows can be exactly matched by using the corresponding dummy variables and a Poisson estimator (ii)

Figure 3.1: Distance elasticities at the State and CFS-area level



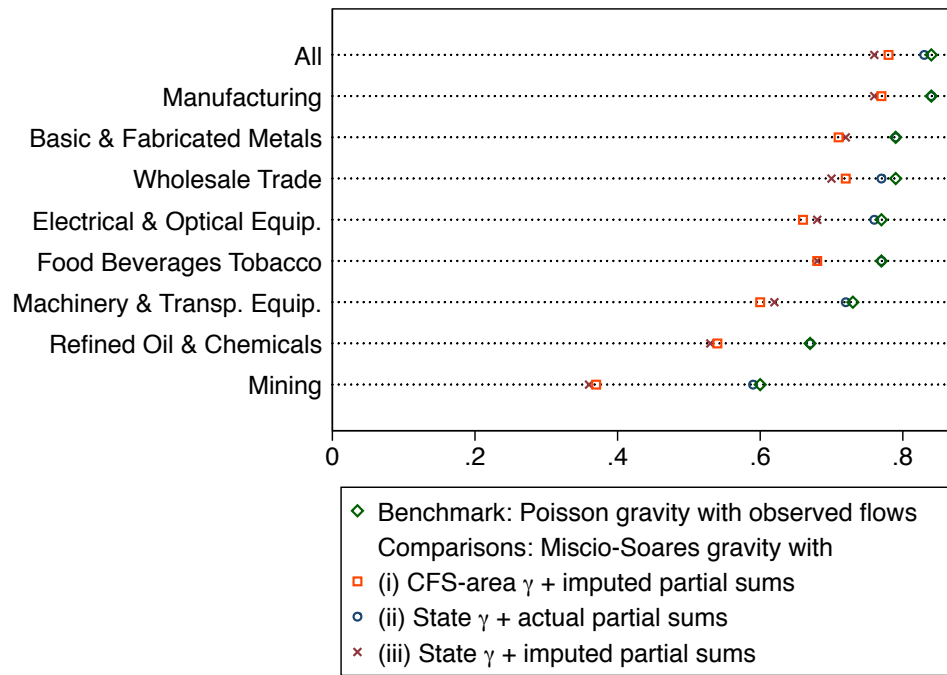
Notes: Distance elasticities from sectoral Poisson gravity regressions using data on trade flows between States and between CFS areas in 2012.

By looking at the benchmark alone, I notice that the ability of gravity models to describe trade patterns varies by industry. One may be surprised to see that the mining sector is not as well described by gravity models as other sectors. However, this is not an artefact of the US data. As I will show in the next section, also when using Brazilian data I find that the lowest correlation is precisely in the mining sector.

Second, I focus on gravity estimation at the CFS-area level and implement the methodology introduced in the previous section. I depart from the standard Poisson estimator by replacing some of its equations with alternative moments that identify the same parameters. This has the effect of relaxing the data requirements. As a reminder, a standard Poisson estimator requires two types of flow data: (i) the partial sums identifying the origin and destination fixed effects, (ii) the matrix of flows identifying the elasticity to

among all possible estimations of gravity equations that belong to the class of general gravity, the ones with a full set of origin and destination fixed effects necessarily provide the highest fit to the data as they have the highest degree of freedom.

Figure 3.2: Correlation between actual and predicted flows at CFS-area level



Notes: Correlations between actual and predicted sectoral trade flows at the CFS-areas level in 2012. Benchmark: Poisson gravity with observed trade flows between CFS areas. Comparisons: Miscio-Soares gravity with distance elasticity γ estimated at the CFS-area level (specification i) or State level (specifications ii & iii) and partial sums (i.e. production and consumption by location) from actual data (specification ii) or imputed (specifications i & iii).

distance. I now focus on each of these, starting with the partial sums. As already noted, the partial sum of all bilateral flows by origin (including internal flows) is the same as production by origin.¹² Similarly, the partial sum by destination is the same as consumption by destination. Therefore, the moment conditions based on partial sums computed from observed flows are numerically equivalent to those computed using *actual* production and consumption data. Therefore, if we have such data we can simply replace the partial sums.¹³ Alternatively, instead of actual consumption and production data by region we may have a model that allows us to compute local production and consumption. For instance, with identical homotetic preferences we can compute local final consumption simply multiplying expenditure shares by local income. I call these *imputed* partial sums because their value may differ from actual production and consumption as it is computed using model-dependent equations. In this exercise, I impute production and consumption by location using the model in [Caliendo *et al.* \(2015\)](#).¹⁴

Next, I can also replace the moment identifying the distance elasticity parameter. In this exercise, I eliminate the need for the last moment condition by calibrating the value of γ to the Poisson elasticities estimated at the State level or at the CFS-area level and shown in figure 3.1. This can be thought of as a way of relaxing the data requirements since to compute the elasticities at the State level I require less detailed flow data than to compute them at the CFS-area level. I relax the two data requirements one at a time, and then

¹²If the set of regions only contains subnational origins and destinations, then the sum of all domestic bilateral flows by origin is the same as production sold domestically by origin, i.e. local production minus (local) foreign exports.

¹³In the original application of this methodology by [Miscio & Soares \(2016\)](#) on commuting data, the partial sums by origin and by destination correspond to workforce by place of residence and by place of work. These can be obtained from, respectively, a standard population and business census. For instance, in the US data County Business Patterns provides employment by location and industry down to the zipcode level, while the microdata from the Population census can be used to compute workforce by industry and by place of residence down to the Public-Use-Microdata-Area level.

¹⁴Refer to chapter 1 for a brief description of the model, or to the original paper for complete details. For the purpose of this chapter, suffices to say that the economy is populated by identical households with homotetic preferences and by intermediate goods producers with a Cobb-Douglas technology that combines labor, structures and material inputs (i.e. other intermediate goods). Therefore, I compute production using data on the local wage bill by industry, I compute intermediate consumption using production and data on intermediate input shares, I compute final consumption using data on local income.

together. For each specification, I estimate all the remaining gravity parameters, that is all the fixed effects, and then I compute the predicted flows using equation 3.1. I assess the goodness of fit by comparing the correlation between actual and predicted flows based on the benchmark Poisson estimator with those obtained using [Miscio & Soares'](#) approach. Figure 3.2 summarises the results.

Overall, the estimates and predictions obtained using [Miscio & Soares'](#) approach are very close to the Poisson benchmark, even in the version with the weakest data requirements, i.e. in the specification with imputed partial sums and where γ is calibrated to the value estimated using CFS data at the State level. In this particular application the largest drop in predictive power is due to the use of model-dependent *imputed* values for local production and consumption, rather than to the use of distance elasticities estimated using data at the State level. This is not surprising since I already showed in figure 3.1 that the choice of the spatial units does not alter the point estimates in a substantial way.

The conclusion from this validation exercise on US data is that the methodology proposed in section 3.2 can be used to compute predicted bilateral trade flows that approximate very well the actual flows. Although the correlation between actual and predicted flows varies by sector and is far from perfect, this methodology is at the moment the only one that can be used to predict sectoral trade flows. In the next section, I apply it to Brazilian data, where we observe sectoral trade flows between States but not between metropolitan areas.

3.3.2 Estimation of Brazilian sectoral domestic trade flows

The only domestic trade data that I came across in the literature is from [de Vasconcelos & de Oliveira \(2006\)](#) and it contains trade flows between most Brazilian States by industry for the year 1999. However, in order to estimate the model in chapter 1 at the metropolitan area level, I require bilateral flows at this spatial scale. In the absence of data at this level of spatial aggregation, I compute an approximation for such flows using the methodology

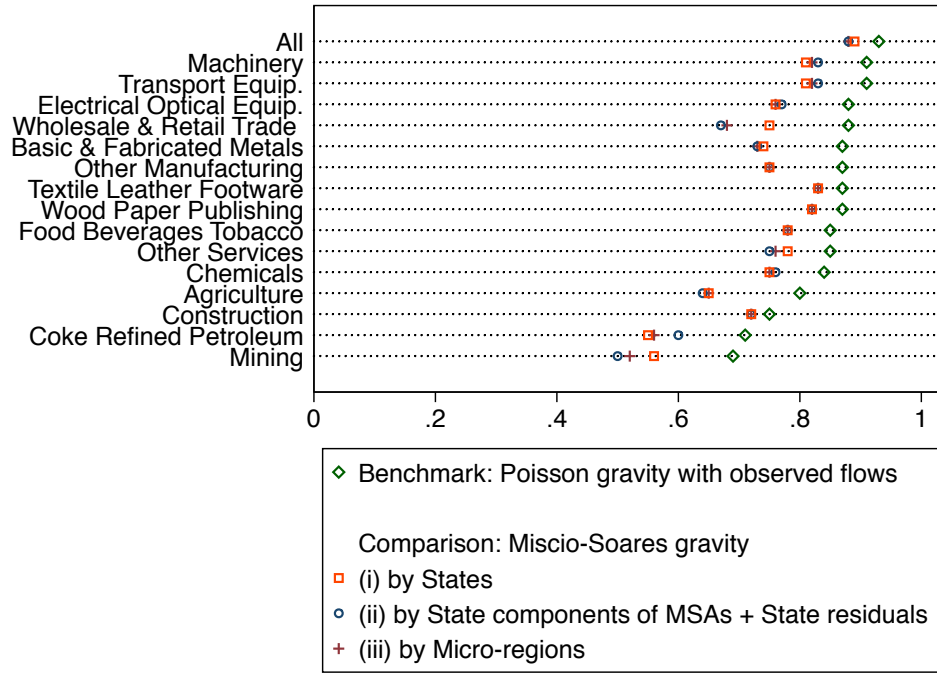
presented in this chapter.

The Brazilian metropolitan areas used in chapter 1 are aggregations of *municípios* defined from the bottom up using census micro-data on workers' residence and workplace location, as explained in section 1.2.1. Since some of these metropolitan areas cross State boundaries, I divide them into State components. The remaining *municípios* are merged into State residuals. Therefore, the spatial units used in this chapter mimic the CFS regions in the US data. The reason for breaking up the metropolitan areas that cross State boundaries is that it enables me to compare predicted flows across levels of aggregation, as I explain next.

In the application to US CFS data presented above, goodness of fit was assessed by comparing observed flows between CFS regions to the predicted counterparts. In the Brazilian context that test is not possible at the metropolitan area level since I only observe flows between States. This problem is similar to the one faced by Monte *et al.* (2015), who only observed trade data between US CFS regions but use their approach to predict flows between counties. As in their paper, I assess goodness of fit in the following way: after computing predicted flows at the more disaggregate level, I aggregate them into predicted flows between States and then compare these aggregate predictions to observed flow data. Hence the need to split the metropolitan areas that cross State boundaries into State components. To ease the exposition, in the rest of this chapter I will refer to these *State components of metropolitan areas and State residuals* simply as *metropolitan areas*.

Specifically, I perform the following estimations. To begin, I estimate a general gravity equation by sector analogous to equation 3.1, with a full set of origin and destination fixed effects at the State level and using a standard Poisson estimator. I compute predicted flows and the correlation between predicted and observed flows in order to establish a benchmark. These correlations are shown in figure 3.3. As in the US data, I notice that the flows in the mining sector are not as well explained by gravity models as the other sectors. Then, as a way to relax the data requirements in steps, I apply Miscio & Soares'

Figure 3.3: Correlation between actual and predicted flows at State level (Brazil)

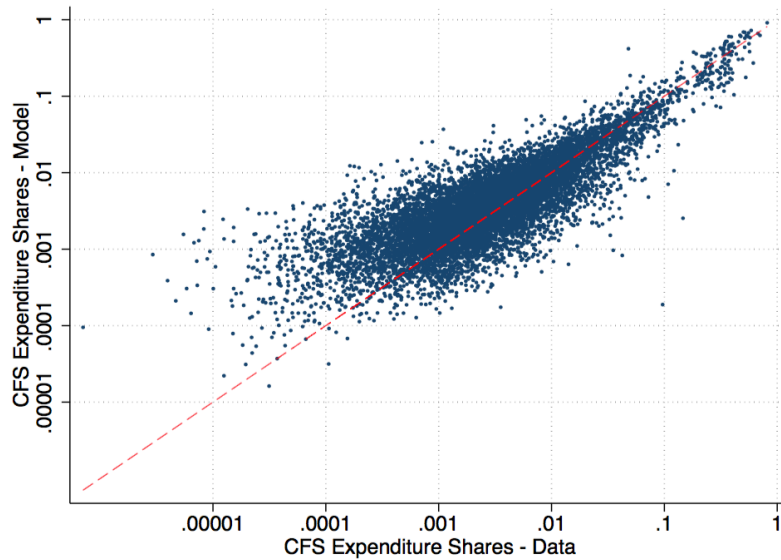


Notes: Correlations between observed flows at the State level and predicted flows aggregated to State level from predictions at lower geographical level (refer to main text for more details). Benchmark: Poisson gravity with observed trade flows between Brazilian States. Comparisons: Miscio-Soares gravity with distance elasticity γ estimated at the State level and partial sums (i.e. local production and consumption) imputed at the level of (i) States; (ii) State components of MSAs and State residuals; (iii) Micro-regions.

methodology to States and then to lower level spatial units. In all cases, the distance elasticity parameters are set to the values estimated in the benchmark Poisson regressions. First, I modify the moment conditions that identify the fixed effects by replacing the partial sums of actual flows between States (i.e. actual production and destination by State) with imputed production and consumption by State. The imputed values are computed using the modeling assumptions from [Caliendo *et al.* \(2015\)](#), like in the US application. Then I estimate the remaining parameters, i.e. the fixed effects, and I compute predicted flows and the correlation with the observed counterparts. Even though I observe a drop in predicted power relative to the Poisson benchmark, these correlations remain high.

Second, I compute imputed production and consumption by metropolitan area and use them in the moment conditions identifying the fixed effects at this lower level of

Figure 3.4: Actual and predicted expenditure shares in [Monte *et al.* \(2015\)](#), sum across sectors



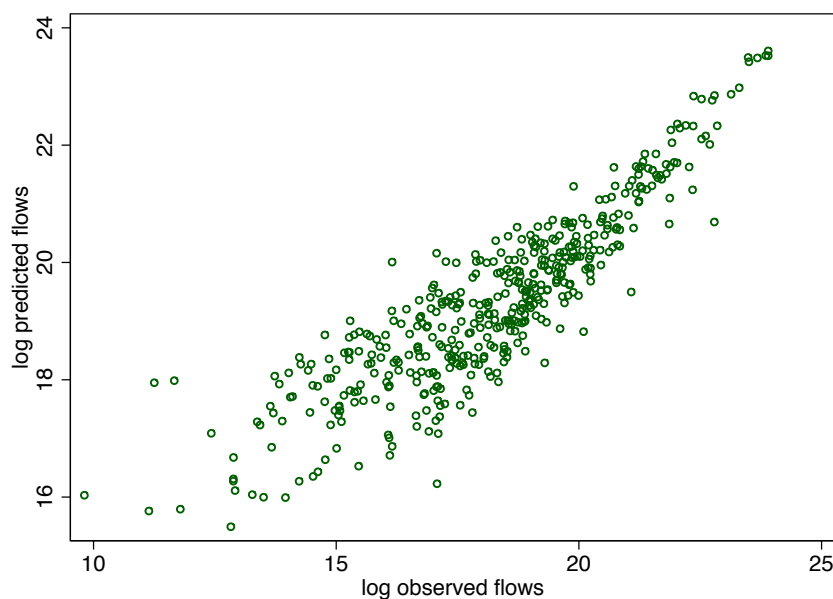
Notes: Observed trade shares between CFS regions (logarithmic scale), summing across all sectors, plotted against predictions computed with [Monte *et al.*](#) 's methodology and aggregated from county level (source: [Monte *et al.* \(2015\)](#), figure 2).

spatial aggregation. The distance elasticities continue to be set to the values estimated on State data. I repeat the same exercise for Micro-regions, which are commonly used in the literature on Brazilian local labor markets and are defined by the Brazilian Institute for Statistics and Geography as a set of municipalities with a common productive structure. I compute the predicted flows and aggregate them to the State level in order to compare them with observed flows. As in the US exercise, the largest drop in predictive power is associated with the replacement of actual partial sums with imputed partial sums. On the other hand, there is hardly any reduction in predictive power when going from a distance elasticity estimated at the same level of spatial aggregation to one estimated at a higher level.

How does this methodology compare to [Monte *et al.*](#) 's in empirical applications? As explained earlier, the gravity equation in [Monte *et al.* \(2015\)](#) only describes aggregate flows, i.e. the total across all sectors, so I will only compare these. Unfortunately, they do not report a goodness of fit statistic to compare actual flows between CFS regions and

predictions aggregated from the county level. However, they display these in graphical format. I paste their figure 2 in figure 3.4. Moreover, they report that the R^2 from a linear fixed effects regression of actual flows between CFS regions is 0.83, which corresponds to a correlation between data and estimation predictions of 0.91. This correlation puts an upper bound on the correlation between actual flows at the CFS-region level and predictions aggregated from the county level. For comparison, figure 3.5 shows Brazilian flows at the State level against predictions based on our methodology and aggregated from metropolitan areas.¹⁵ The correlation between the flows displayed in figure 3.5 is 0.88 and the linear regression on State level data analogous to the one in *Monte et al.* yields an R^2 of 0.90, which corresponds to a correlation of 0.95. Therefore, based on our respective exercises the goodness of fit of the two methodologies seems very comparable. However, as already explained, the big difference between the two methodologies is that ours can also be used for sectoral trade flows.

Figure 3.5: Actual and predicted flows, sum across sectors



Notes: Observed flows between Brazilian States (in logarithms), summing across all sectors, plotted against predictions computed with *Miscio & Soares (2016)*'s methodology and aggregated from State components of metropolitan areas and State residuals.

¹⁵Notice that *Monte et al. (2015)* report expenditure shares on a log scale while I report log flows. Given that my estimator satisfies the adding-up property, log expenditure shares are simply a linear transformation of log flows. Therefore, our plots are comparable even though the displayed units are different.

3.4 CONCLUSION

In this chapter I showed how to use the methodology by [Miscio & Soares \(2016\)](#) to estimate the parameters of a generic gravity equation without using bilateral flow data. In particular, I showed how to interpret the first order conditions of a standard Poisson estimator as the moments of a General Method of Moments estimator, and how to modify these moments so that they do not require the full matrix of bilateral flow data. I discussed alternative approaches found in the recent literature and how they compare with the one presented here.

I applied this methodology to US Commodity Flow Survey data in order to compare observed flows between CFS regions by sector to predicted counterparts. The results suggest that by using this methodology it is possible to obtain predictions that are highly correlated with the observed flows. Then, I turned to Brazilian domestic trade data to compute predicted flows between metropolitan areas, for which there is no bilateral trade data. I conclude by arguing that while the predictive power of this methodology is comparable to one of the alternatives found in the literature, this is the only one that can be used to also estimate bilateral flows by sector.

Bibliography

- Ahlfeldt, Gabriel M., Redding, Stephen J., Sturm, Daniel M., & Wolf, Nikolaus. 2015. The Economics of Density: Evidence From the Berlin Wall. *Econometrica*, **83**(6), 2127–2189.
- Albouy, David. 2009. The Unequal Geographic Burden of Federal Taxation. *Journal of Political Economy*, **117**(4), pp. 635–667.
- Amiti, Mary, & Cameron, Lisa. 2007. Economic Geography and Wages. *Review of Economics and Statistics*, **89**(1), pp. 15–29.
- Arellano, Manuel, & Bond, Stephen. 1991. Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations. *The Review of Economic Studies*, **58**(2), 277–297.
- Armington, Paul S. 1969 (March). *A Theory of Demand for Products Distinguished by Place of Production*. IMF Staff Papers.
- Artuc, Erhan, Chaudhuri, Shubham, & McLaren, John. 2010. Trade Shocks and Labor Adjustment: A Structural Empirical Approach. *American Economic Review*, **100**(3), 1008–10045.
- Arvis, Jean-Francois, & Shepherd, Ben. 2013. The Poisson quasi-maximum likelihood estimator: a solution to the “adding up” problem in gravity models. *Applied Economic Letters*, **20**, 515–519.
- Autor, David H, Dorn, David, & Hanson, Gordon H. 2013. The China Syndrome: Local Labor Market Effects of Import Competition in the United States. *American Economic Review*, **103**(6), 2121–2168.

- Autor, David H., Dorn, David, Hanson, Gordon H., & Song, Jae. 2014. Adjustment to Trade: Worker Level Evidence. *Quarterly Journal of Economics*, **129**(4), 1799–1860.
- Baines, Dudley. 1985. *Migration in a Mature Economy*. Cambridge, UK: Cambridge University Press.
- Baines, Dudley. 1994. Population, Migration and Regional Development, 1870-1939. In: Floud, R, & McCloskey, D (eds), *The Economic History of Britain Since 1700*, second edition edn. Cambridge University Press.
- Balsvik, Ragnhild. 2011. Is labor mobility a channel for spillovers from multinationals? Evidence from Norwegian manufacturing. *Review of Economics and Statistics*, **93**(1), 285–297.
- Barrios, Thomas, Diamond, Rebecca, Imbens, Guido, & Kolesar, Michal and. 2012. Clustering, Spatial Correlations and Randomized Inference. *Journal of the American Statistical Association*, **107**(498), 578–591.
- Barro, Robert J. 2012 (August). *Convergence and Modernization Revisited*. NBER Working Paper No. 18295.
- Barro, Robert J., & Sala-i Martin, Xavier. 1992. Convergence. *Journal of Political Economy*, **100**(2), pp. 223–251.
- Bartik, Timothy J. 1991. *Who Benefits from State and Local Economic Development Policies?* Kalamazoo, MI: W.E. Upjohn Institute for Employment Research.
- Baum-Snow, Nathaniel, & Ferreira, Fernando. 2015. Causal Inference in Urban and Regional Economics. In: Duranton, Gilles, Henderson, J. Vernon, & Strange, William C. (eds), *Handbook of Regional Urban Economics*. Handbook of Regional and Urban Economics, vol. 5.
- Bertrand, Marianne, Duflo, Esther, & Mullainathan, Sendhil. 2004. How Much Should We Trust Differences-in-Differences Estimates? *The Quarterly Journal of Economics*, **119**(1), pp. 249–275.
- Bishop, Kelly C. 2012. *A Dynamic Model of Location Choice and Hedonic Valuation*. Unpub-

lished Manuscript.

- Bleakley, Hoyt, & Lin, Jeffrey. 2012. Portage: Path Dependence and Increasing Returns in U.S. History. *Quarterly Journal of Economics*, **127**(May), 587–644.
- Blundell, Richard, & Bond, Stephen. 1998. Initial Conditions and Moment Restrictions in Dynamic Panel Data Models. *Journal of Econometrics*, **87**(1), 115 – 143.
- Caliendo, Lorenzo, & Parro, Fernando. 2015. Estimates of the Trade and Welfare Effects of NAFTA. *Review of Economic Studies*, **82**(1), 1–44.
- Caliendo, Lorenzo, Dvorkin, Maximiliano, & Parro, Fernando. 2015 (May). *Trade and Labor Market Dynamics*. NBER Working Paper No. 21149.
- Cameron, A. C., Gelbach, J.B., & Milller, D. L. 2011. Robust Inference with Multi-Way Clustering. *Journal of Business and Economic Statistics*, **29**(2), 238–249.
- Caselli, Francesco, Esquivel, Gerardo, & Lefort, Fernando. 1996. Reopening the Convergence Debate: A New Look at Cross-Country Growth Empirics. *Journal of Economic Growth*, **1**(3), pp. 363–389.
- Chaney, Thomas. 2008. Distorted Gravity: The Intensive and Extensive Margins of International Trade. *The American Economic Review*, **98**(4), pp. 1707–1721.
- Chay, Kenneth Y., & Greenstone, Michael. 2005. Does Air Quality Matter? Evidence from the Housing Market. *Journal of Political Economy*, **113**(2), pp. 376–424.
- Chinitz, Benjamin. 1961. Contrasts in Agglomeration: New York and Pittsburgh. *The American Economic Review*, **51**(2), pp. 279–289.
- Combes, Pierre-Philippe. 2000. Economic Structure and Local Growth: France, 1984-1993. *Journal of Urban Economics*, **47**(3), 329 – 355.
- Combes, Pierre-Philippe, & Gobillon, Laurent. 2015. The Empirics of Agglomeration Economies. *Pages pp. 247–348 of: Duranton, Gilles, Henderson, Vernon, & Strange, William (eds), Handbook of Urban and Regional Economics*, vol. 5. Elsevier.
- Combes, Pierre-Philippe, Duranton, Gilles, & Gobillon, Laurent. 2008. Spatial Wage Disparities: Sorting Matters!. *Journal of Urban Economics*, **63**(2), 723 – 742.

- Combes, Pierre-Philippe, Duranton, Gilles, & Gobillon, Laurent. 2011. The Identification of Agglomeration Economies. *Journal of Economic Geography*, **11**(2), 253 – 266.
- Combes, Pierre-Philippe, Duranton, Gilles, Gobillon, Laurent, Puga, Diego, & Roux, Sebastien. 2012. The Productivity Advantages of Large Cities: Distinguishing Agglomeration from Firm Selection. *Econometrica*, **80**(6), pp. 2543–2594.
- Conley, Timothy G. 1999. GMM Estimation with Cross Sectional Dependence. *Journal of Econometrics*, **92**(1), 1 – 45.
- Costa, Francisco, Garred, Jason, & Pessoa, João P. Forthcoming. Winners and Losers from a Commodities-for-Manufactures Trade Boom. *Journal of International Economics*.
- Crafts, Nicholas, & Mulatu, Abay. 2006. How Did the Location of Industry Respond to Falling Transport Costs in Britain before World War I? *The Journal of Economic History*, **66**(3), pp. 575–607.
- Davis, Donald R, & Weinstein, David E. 2002. Bones, bombs, and break points: The geography of economic activity. *American Economic Review*, **92**(5), 1269–1289.
- de Vasconcelos, Jose R., & de Oliveira, Marcio A. 2006 (February). *IPEA Discussion Paper no. 1159*.
- Dekle, Robert. 2002. Industrial Concentration and Regional Growth: Evidence from the Prefectures. *The Review of Economics and Statistics*, **84**(2), pp. 310–315.
- Desmet, Klaus, & Rappaport, Jordan. Forthcoming. The Settlement of the United States, 1800-2000: The Long Transition Towards Gibrat's Law. *Journal of Urban Economics*.
- Desmet, Klaus, & Rossi-Hansberg, Esteban. 2014. Spatial Development. *American Economic Review*, **104**(4), 1211–43.
- Dhyne, Emmanuel, Magerman, Glenn, & Rubínoví, Stela. 2015. *The Belgian production network 2002-2012*. National Bank of Belgium Working Paper No. 288.
- Diamond, Rebecca. Forthcoming. The Determinants and Welfare Implications of US Workers' Diverging Location Choices by Skill: 1980-2000. *American Economic Review*.
- Dittmar, Jeremiah E. 2011 (August). *Cities, Markets, and Growth: The Emergence of Zipf's*

Law. Working Paper.

- Dix-Carneiro, Rafael. 2014. Trade Liberalization and Labor Market Dynamics. *Econometrica*, **82**(3), 825–855.
- Dix-Carneiro, Rafael, & Kovak, Brian. 2015 (January). *Trade Reform and Regional Dynamics: Evidence from 25 Years of Brazilian Matched Employer-Employee Data*. NBER Working Paper No. 20908.
- Dumais, Guy, Ellison, Glenn, & Glaeser, Edward L. 1997 (November). *Geographic Concentration as a Dynamic Process*. NBER Working Paper No. 6720.
- Dumais, Guy, Ellison, Glenn, & Glaeser, Edward L. 2002. Geographic Concentration as a Dynamic Process. *The Review of Economics and Statistics*, **84**(2), pp. 193–204.
- Duranton, Gilles. 2013 (April). *Delineating metropolitan areas: Measuring spatial labour market networks through commuting patterns*. Working paper.
- Duranton, Gilles, Morrow, Peter M., & Turner, Matthew A. 2014. Roads and Trade: Evidence from the US. *The Review of Economic Studies*, **81**(2), 681–724.
- Eaton, Jonathan, & Kortum, Samuel. 2002. Technology, Geography, and Trade. *Econometrica*, **70**(5), pp. 1741–1779.
- Ellison, G., Glaeser, E., & Kerr, W. 2010. What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns. *American Economic Review*, **100**(3), pp. 1195–1213.
- Ellison, Glenn, & Glaeser, Edward L. 1997. Geographic concentration in US manufacturing industries: A dartboard approach. *Journal of Political Economy*, **105**(5), 889–927.
- Faggio, Guilia, Silva, Olmo, & Strange, William. Forthcoming. Heterogeneous Agglomeration. *Review of Economics and Statistics*, March.
- Fally, Thibault. 2015. Structural gravity and fixed effects. *Journal of International Economics*, **97**, 76–85.
- Glaeser, Edward L., & Mare, David C. 2001. Cities and Skills. *Journal of Labor Economics*, **19**(2), pp. 316–342.
- Glaeser, Edward L, Kallal, Hedi D, Scheinkman, Jose A, & Shleifer, Andrei. 1992. Growth

- in Cities. *Journal of Political Economy*, **100**(6), 1126–1152.
- Greenstone, Michael, & Moretti, Enrico. 2003 (July). *Bidding for Industrial Plants: Does Winning a 'Million Dollar Plant' Increase Welfare?* NBER Working Paper No. 9844.
- Greenstone, Michael, Hornbeck, Richard, & Moretti, Enrico. 2010. Identifying Agglomeration Spillovers: Evidence from Million Dollar Plants. *Journal of Political Economy*, **118**(3), pp. 536–598.
- Gyourko, Joseph, Saiz, Albert, & Summers, Anita. 2008. A new measure of the local regulatory environment for housing markets: The Wharton Residential Land Use Regulatory Index. *Urban Studies*, **45**(3), 693–729.
- Hanson, Gordon H. 2005. Market potential, increasing returns and geographic concentration. *Journal of International Economics*, **67**(1), 1 – 24.
- Hauk Jr., William R., & Wacziarg, Romain. 2009. A Monte Carlo Study of Growth Regressions. *Journal of Economic Growth*, **14**(2), pp. 103–147.
- Head, Keith, & Mayer, Thierry. 2000. Non-Europe: the magnitude and causes of market fragmentation in the EU. *Review of World Economics*, 284–314.
- Head, Keith, & Mayer, Thierry. 2014. Gravity Equations: Workhorse, Toolkit and Cookbook. In: Gopinath, G, Helpman, E., & Rogoff, K. (eds), *Handbook of International Economics*, vol. 4. Elsevier.
- Henderson, J. Vernon. 1996. Effects of Air Quality Regulation. *The American Economic Review*, **86**(4), pp. 789–813.
- Henderson, J Vernon. 1997. Externalities and Industrial Development. *Journal of Urban Economics*, **42**(3), 449–470.
- Henderson, V, Kuncoro, A, & Turner, M. 1995. Industrial Development in Cities. *Journal of Political Economy*, **103**(5), 1067–1090.
- Horrell, Sara, Humphries, Jane, & Weale, Martin. 1994. An Input-Output Table for 1841. *Economic History Review*, **47**(3), 545–566.
- Hurwicz, Leonid. 1950. Least-Squares Bias in Time Series. In: Koopmans, Tjalling C. (ed),

- Statistical Inference in Dynamic Economic Models*. New York: Wiley.
- Jacks, David S., Meissner, Christopher M., & Novy, Dennis. 2008. Trade Costs, 1870-2000. *The American Economic Review*, **98**(2), pp. 529–534.
- Jacobs, Jane. 1969. *The Economy of Cities*. New York: Vintage Books.
- Javorcik, Beata S. 2004. Does Foreign Direct Investment Increase the Productivity of Domestic Firms? In Search of Spillovers through Backward Linkages. *The American Economic Review*, **94**(3), 605–627.
- Jones, Ronald W. 1975. Income Distribution and Effective Protection in a Multicommodity Trade Model. *Journal of Economic Theory*, **11**(1), 1 – 15.
- Kennan, John, & Walker, James R. 2011. The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, **79**(1), 211–251.
- Kim, Sukkoo. 1995. Expansion of Markets and the Geographic Distribution of Economic Activities - The Trends in US Regional Manufacturing Structure, 1860-1987. *Quarterly Journal of Economics*, **110**(4), 881–908.
- Kleibergen, Frank, & Paap, Richard. 2006. Generalized Reduced Rank Tests Using the Singular Value Decomposition. *Journal of Econometrics*, **133**, 97–126.
- Kleibergen, F. and Schaffer, M.E. 2010. *Rankest: Module for Testing the Rank of a Matrix using the Kleibergen-Paap rk Statistic*. Available at: <http://ideas.repec.org/c/boc/bocode/s456865.html>.
- Kline, Patrick, & Moretti, Enrico. 2013. Local Economic Development, Agglomeration Economies, and the Big Push: 100 Years of Evidence from the Tennessee Valley Authority. *Quarterly Journal of Economics*.
- Kovak, Brian K. 2013. Regional Effects of Trade Reform: What is the Correct Measure of Liberalization? *American Economic Review*, **103**(5), 1960–1976.
- Krugman, Paul. 1991. Increasing Returns and Economic Geography. *Journal of Political Economy*, **99**(3), pp. 483–499.
- Kugler, Maurice. 2006. Spillovers from Foreign Direct Investment: Within or Between

- Industries? *Journal of Development Economics*, **80**(2), 444–477.
- Lee, Jamie. 2015 (January). *Measuring Agglomeration: Products, People and Ideas in U.S. Manufacturing, 1880-1990*. Mimeo.
- Long, Jason, & Ferrie, Joseph. 2003. Labour Mobility. In: Mokyr, Joel (ed), *Oxford Encyclopedia of Economic History*. New York: Oxford University Press.
- Long, Jason, & Ferrie, Joseph. 2004 (March). *Geographic and Occupational Mobility in Britain and the U.S., 1850-1881*. Working Paper.
- Lopez, R, & Sudekum, J. 2009. Vertical Industry Relations, Spillovers, and Productivity: Evidence from Chilean Plants. *Journal of Regional Science*, **49**(4), pp. 721–747.
- Lucas, Robert E. 1988. On the Mechanics of Economic Development. *Journal of Monetary Economics*, **22**(1), 3–42.
- Marshall, Alfred. 1890. *Principles of Economics*. New York: Macmillan and Co.
- Melitz, MJ. 2003. The impact of trade on intra-industry reallocations and aggregate industry productivity. *Econometrica*, **71**(6), 1695–1725.
- Michaels, Guy, Rauch, Ferdinand, & Redding, Stephen. 2013 (January). *Task Specialization in U.S. Cities from 1880-2000*. NBER Working Paper No. 18715.
- Miscio, Antonio, & Soares, Ilton G. 2016. Predicting Flows with Incomplete Data: Example for commuting with applications. *Unpublished manuscript*.
- Monte, Ferdinando, Redding, Stephen J., & Rossi-Hansberg, Esteban. 2015 (November). *Commuting, Migration and Local Employment Elasticities*. NBER Working Paper No. 21706.
- Nickell, Stephen. 1981. Biases in Dynamic Models with Fixed Effects. *Econometrica*, **49**(6), pp. 1417–1426.
- Notowidigdo, Matthew J. 2013 (March). *The Incidence of Local Labor Demand Shocks*. Working Paper.
- Poole, Jennifer P. 2013. Knowledge Transfers from Multinational to Domestic Firms: Evidence from Worker Mobility. *The Review of Economics and Statistics*, **95**(2).

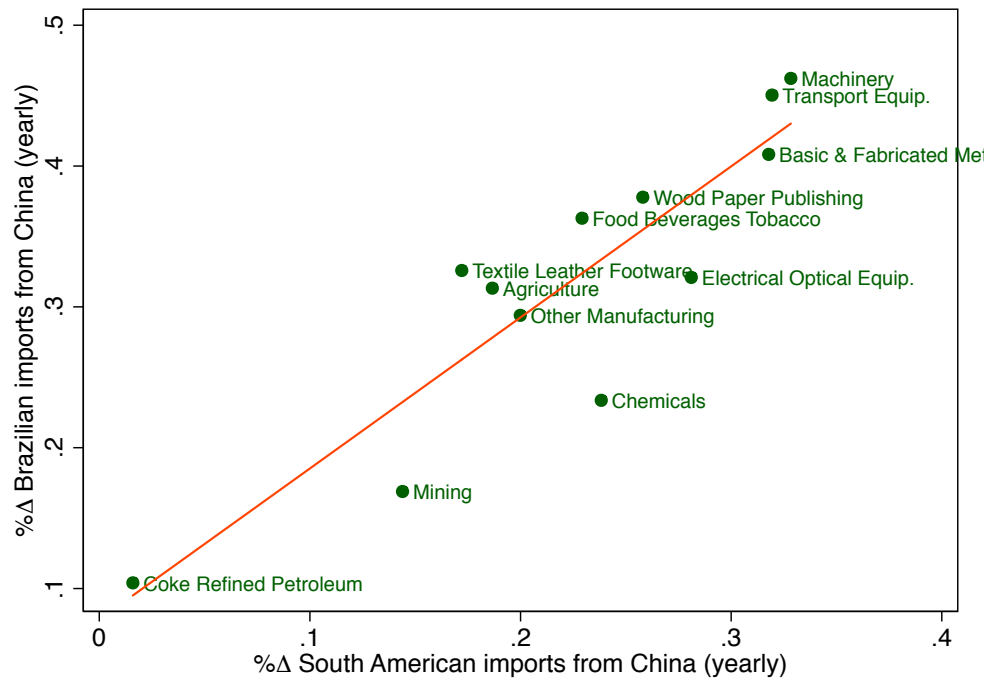
- Rosenthal, S., & Strange, W. 2004. Evidence on the Nature and Sources of Agglomeration Economies. In: Henderson, J.V., & Thisse, J.F. (eds), *Handbook of Regional and Urban Economics*. Elsevier.
- Rosenthal, Stuart S., & Strange, William C. 2001. The Determinants of Agglomeration. *Journal of Urban Economics*, **50**(2), 191–229.
- Rosenthal, Stuart S., & Strange, William C. 2003. Geography, Industrial Organization, and Agglomeration. *The Review of Economics and Statistics*, **85**(2), pp. 377–393.
- Schmutz, Benoit, & Sidibe, Modibo. 2015. *Job Search and Migration in a System of Cities*. Unpublished Manuscript.
- Silva, Santos J.M.C., & Tenreyro, Silvana. 2006. The log of gravity. *Review of Economics and Statistics*, **88**, 641–658.
- Sizer, Molly, & Tolbert, Charles M. 1996. *U.S. Commuting Zones and Labor Market Areas: a 1990 Update*. U.S. Department of Agriculture Staff Paper, AGES-9614.
- Thomas, Mark. 1987. *An Input-Output Approach to the British Economy, 1890-1914*. Ph.D. thesis, Oxford University.
- Thompson, Samuel B. 2011. Simple Formulas for Standard Errors that Cluster by both Firm and Time. *Journal of Financial Economics*, **99**(1), 1 – 10.
- Topalova, Petia. 2007. *Trade Liberalization, Poverty and Inequality: Evidence from Indian Districts*. University of Chicago Press. Chap. Globalization and Poverty, pages 291–336.
- Yagan, Danny. 2014 (January). *Moving to Opportunity? Migratory Insurance Over the Great Recession*. Working Paper.

Appendix

APPENDIX TO CHAPTER 1

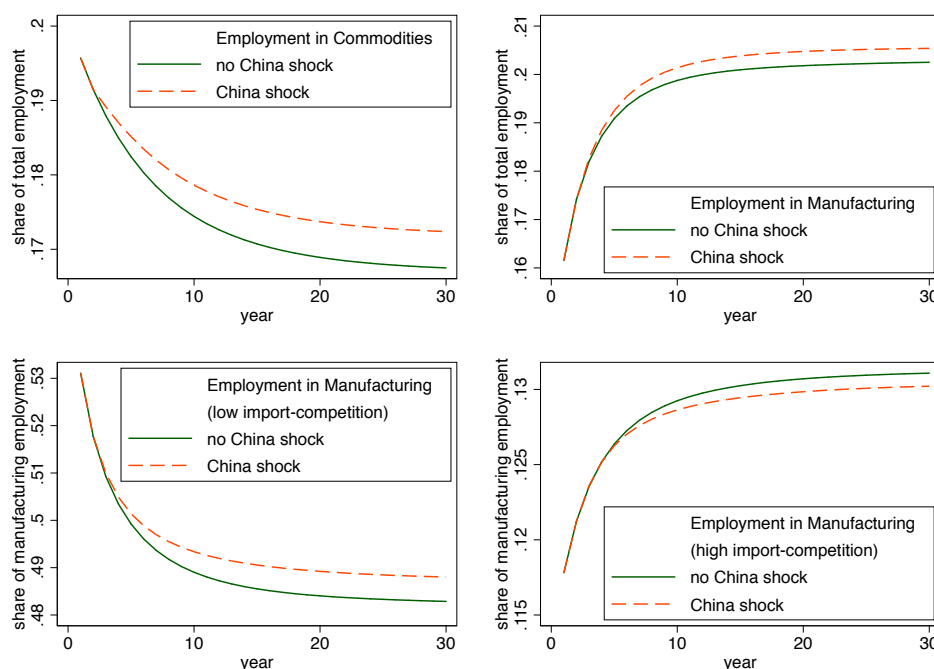
Calibration of Chinese productivity changes and model predictions

Figure 3.6: Growth in real imports from China to Brazil and to other South American countries



Notes: Annualized percentage change in real imports from China to Brazil and to other South American countries in 2000-2010. Correlation = 0.89. Data: BACI.

Figure 3.7: Employment shares predicted by the model with and without China shock



Notes: Evolution of sectoral employment shares predicted by the model using data at the metropolitan area level. Commodity sector includes Agriculture and Mining. Low and high import competition manufacturing includes the three sectors with lowest and highest change in Brazilian imports from China per worker, as shown in figure 1.6.

APPENDIX TO CHAPTER 2

Theory appendix

This appendix explores several additional factors that are not included in the model provided in the main text. We begin by discussing the implications of allowing variation in the costs of innovation or firm entry across industries. Next, we consider allowing production function parameters to vary across industries. We then look at incorporating capital into the model. Finally, we consider the implications of including intermediate goods and the closely related issue of incorporating trade costs into the model.

Variation in Industry Innovation or Entry Cost Parameters

Suppose that the innovation costs or entry costs are allowed to vary by industry, so

that we have F_i and C_i . This will affect the rate of firm entry and R&D, which will affect the size of city-industry employment. However, as long as these cost parameters are fixed over time and both denominated in the same units (labor), they will be differenced out when we obtain the main regression specification. Thus, our empirical approach will be robust to this modification. Note that holding these parameters fixed over time does not imply that the costs of entry or innovation is fixed over time, since that cost will also depend on the wage, which will vary over time and across locations. However, it does imply that the relative cost of entry and R&D is constant over time and across locations, even if it varies across industries.

Variation in Industry Production Function Parameters

Suppose that we allow the production function parameters to vary by industry, so that they are now all indexed by i . In this case, the coefficient on the spillover term from industry k to industry i , which in our baseline model is τ_{ki}/β is now given by τ_{ki}/β_i . What this tells us is that the way in which employment in industry k is translated into employment growth in industry i will now depend on the importance of local industry-specific resources in the production function. This is because local resources are the factors that tie industries to particular locations. Industries in which local resources are relatively unimportant (low β) should exhibit large estimated spillover coefficients because employment will be more able to respond to changing technology levels by shifting across locations.

When we estimate single cross-industry spillover terms we will be estimating the average impact across industries with potentially varying β parameters. It is also possible to estimate industry-specific spillover benefit terms in our framework, which is done in [Appendix 2.3](#).

Incorporating Capital

Suppose that we incorporate capital as an input into production, so that the production function is,

$$y_{icft} = a_{icft} L_{icft}^{\alpha_1} H_{icft}^{\alpha_2} K_{icft}^{\alpha_3} R_{icft}^{\beta} E_{icft}^{1-\alpha_1-\alpha_2-\beta}.$$

If there is a national capital market then this implies a national price of capital, which we denote κ_t . In this case, introducing capital into the model would simply add an additional time-varying national factor to the estimating equation. This will be absorbed by the fixed effects and would not affect our results.

Alternatively, it may be the case that capital markets are more local. In this case, the price of capital will be κ_{ct} . When industries share a common set of production function parameters this will affect all industries in a similar way. As a result, it will be absorbed into the city-time effects and will not impact our results.

However, if we also allow industries to be more or less capital intensive, then variation in the local price of capital may have heterogeneous effects on city-industry growth. Under these circumstances, the source of capital will become important. If capital is related to city size, for example because capital availability depends on total local savings which scales with city size, then this will introduce an industry-specific city size effect. It is possible to incorporate industry-specific city size effects into our regression framework. We have experimented with doing so and it does not substantially alter our results. In order for capital to be one channel behind our within and cross-industry spillover estimates, we need three things to be true: (1) capital is local, (2) capital intensity varies across industries, and (3) the accumulation of savings depends on the local composition of industries. If all of these factors are in place, then local capital channels may be a dynamic agglomeration force in our setting. However, capital was fairly mobile across regions in Britain during the period we study, suggesting that local capital accumulation is unlikely to be behind our results, though this may have been a more important factor during the earlier first Industrial Revolution period.

Intermediate Goods and Trade Costs

Next, we discuss the implications of incorporating intermediate goods and trade costs into the theory. We begin by introducing intermediate inputs while maintaining the assumption of free trade. We then consider the implications of allowing non-zero trade costs.

Suppose that each firm uses a basket of intermediate inputs denoted by I_{icft} in production, with a production function parameter φ . Let the set of intermediate inputs used in production vary across industries, but for simplicity, we assume that within an industry all firms use these inputs in fixed proportions. Let Z be an input-output matrix with elements z_{ij} such that I_{it} units of intermediate input to industry i require $I_{it}z_{ij}$ units of output from industry j (i.e., the production function for intermediates is Leontief). Total intermediate demand for the output from industry j is then $x_{jt}^{IO} = \sum_i I_{it}z_{ij}$. With costless trade, each industry will face a national intermediate good input price in each period, which we denote d_{it} . The resulting firm optimization problem in period three is,

$$\max_{L_{icft}, R_{icft}} p_{it} a_{icft} L_{icft}^{\alpha_1} H_{icft}^{\alpha_2} I_{icft}^{\varphi} R_{icft}^{\beta} - w_{ct} L_{icft} - q_{ct} H_{icft} - d_{it} I_{icft} - r_{ict} R_{icft}$$

with $1 - \alpha_1 - \alpha_2 - \varphi - \beta > 0$

With free trade, this will yield a regression specification that is very similar to the one obtained in the main text:

$$\begin{aligned}
\ln(L_{ict}) - \ln(L_{ict-1}) &= \left(\frac{1}{\beta} \right) \left[\tau_{ii} \ln(L_{ict}) + \sum_{k \neq i} \tau_{ki} \ln(L_{kct}) \right. \\
&- \varphi \left[\ln(d_{it}) - \ln(d_{it-1}) \right] \\
&+ \left[\ln(p_{it}) - \ln(p_{it-1}) \right] + \xi_{it} \\
&- \left[v(\alpha_2 - 1) - \alpha_2 \right] \left[\ln(q_{ct}) - \ln(q_{ct-1}) \right] + \psi_{ct} \\
&\left. + (\alpha_2 - 1) \left[\ln(\bar{v}_t^*) - \ln(\bar{v}_{t-1}^*) \right] + \epsilon_{ict} \right]
\end{aligned} \tag{3.5}$$

Thus, under the assumption of free trade across locations, the introduction of intermediate inputs will not impact our results because the impact of changing input prices will be absorbed in the time-varying industry effects.

We can use Eq. 3.5 to explore the impact of introducing trade costs into the model in a partial equilibrium way. Allowing non-zero trade costs will affect this equation in two ways. First, output prices will vary at the local level, so p_{it} will become p_{ict} . Second, intermediate input prices will also vary locally, so d_{it} becomes d_{ict} . With trade costs, both the input and the output prices faced by firms in industry i can vary across cities.

To consider the impact of trade costs, suppose for now that we turn off all spillover channels, so $S_{ict} = 0$ and,

$$\begin{aligned}
\ln(L_{ict}) - \ln(L_{ict-1}) &- \varphi \left[\ln(d_{it}) - \ln(d_{it-1}) \right] \\
&+ \left[\ln(p_{it}) - \ln(p_{it-1}) \right] \\
&- \left[v(\alpha_2 - 1) - \alpha_2 \right] \left[\ln(q_{ct}) - \ln(q_{ct-1}) \right] \\
&+ (\alpha_2 - 1) \left[\ln(\bar{v}_t^*) - \ln(\bar{v}_{t-1}^*) \right] + \epsilon_{ict}
\end{aligned} \tag{3.6}$$

Now, focusing on the input prices side, suppose that there are two cities, A and B, and

that City A has more industry i suppliers than city B so that the cost of intermediate inputs to industry i is lower in City A than in City B. This implies that employment in industry i will be larger in City A than in City B in some initial period: this is static agglomeration a la [Krugman \(1991\)](#). A similar effect can be generated through output price channels. However, as we roll the model forward, Equation 3.6 shows that, absent other changes, industry i will *not* grow faster in City A than in City B. In the absence of other effects, input-output connections alone cannot act as a dynamic agglomeration force.

Where input-output connections can generate dynamic agglomeration patterns is by transmitting the effects of other changes, such as falling transport costs. However, falling trade costs cannot be a sustained force of dynamic agglomeration since trade costs are bounded below by zero. Moreover, trade costs were fairly stable over at least part of the period we study, while urbanization continued apace.¹⁶ This pattern suggests that input-output connections and trade costs can be an important static agglomeration force, but these forces are unlikely to generate the dynamic agglomeration patterns studied here.

In a world of static inter-industry agglomeration forces, the growth in industry i must be driven by growth in industry j , rather than the level of industry j . But this raises questions about the causes of the initial growth in industry j . Ultimately, a world of static agglomeration forces is a world of exogenous city-industry growth. In contrast, dynamic agglomeration offers an explanation for city industry growth, just as endogenous growth theory offers an explanation for aggregate growth.

¹⁶[Crafts & Mulatu \(2006\)](#) conclude that, “falling transport costs had only weak effects on the location of industry in the period 1870 to 1911.” [Jacks *et al.* \(2008\)](#) find a rapid fall in external trade costs prior to 1880, with a much slower decline thereafter.

Data appendix

Table 3.1: Cities in the primary analysis database

City	Population in 1851	Working population in 1851	Workers in analysis industries in 1851
Bath	54,240	27,623	22,836
Birmingham	232,841	111,992	94,188
Blackburn	46,536	26,211	24,279
Bolton	61,171	31,211	28,576
Bradford	103,778	58,408	54,685
Brighton	69,673	32,949	27,151
Bristol	137,328	64,025	53,361
Derby	40,609	19,299	16,354
Gateshead	25,568	10,003	8,373
Halifax	33,582	18,058	16,171
Huddersfield	30,880	13,922	12,132
Hull	84,690	36,983	30,810
Ipswich	32,914	14,660	11,745
Leeds	172,270	83,570	73,696
Leicester	60,584	31,140	28,097
Liverpool	375,955	165,300	137,759
London	2,362,236	1,088,285	880,602
Manchester	401,321	204,688	183,406
Newcastle-upon-Tyne	87,784	38,564	32,136
Northampton	26,657	13,626	11,839
Norwich	68,195	34,114	29,032
Nottingham	57,407	33,967	30,538
Oldham	72,357	38,853	35,911
Portsmouth	72,096	31,345	18,536
Preston	69,542	36,864	32,696
Sheffield	135,310	58,551	51,092
South shields	28,974	11,114	9,895
Southampton	35,305	14,999	11,845
Stockport	53,835	30,128	27,676
Sunderland	63,897	24,779	21,302
Wolverhampton	49,985	22,727	19,495

Figure 3.8: Map showing the location of cities in the analysis database



Table 3.2: Industries in the primary analysis database with 1851 employment

Manufacturing		Services and Professional	
Chemicals & drugs	11,501	Professionals*	40,733
Clothing, shoes, etc.	328,669	General services	458,808
Instruments & jewelry*	31,048	Merchant, agent, accountant, etc.	62,564
Earthenware & bricks	19,580	Messenger, porter, etc.	72,155
Leather & hair goods	26,737	Shopkeeper, salesmen, etc.	27,232
Metal & Machines	167,052		
Oil, soap, etc.	12,188		
Paper and publishing	42,578	Transportation services	
Shipbuilding	14,498	Railway transport	10,699
Textiles	315,646	Road transport	40,106
Vehicles	9,021	Sea & canal transport	66,360
Wood & furniture	69,648		
Food, etc.		Others industries	
Food processing	113,610	Construction	137,056
Spiritous drinks, etc.	8,179	Mining	18,413
Tobacconists*	3,298	Water & gas services	3,914

Industries marked with a * are available in the database but are not used in the baseline analysis because they cannot be linked to categories in the 1907 British input-output table.

Table 3.3: Industry agglomeration patterns based on the Ellison & Glaeser index

Industry	1851	1861	1871	1881	1891	1901	1911
Textiles	0.166	0.178	0.182	0.195	0.182	0.159	0.150
Mining	0.143	0.167	0.150	0.199	0.213	0.200	0.197
Metal & machines	0.094	0.089	0.086	0.093	0.087	0.079	0.088
Shipbuilding	0.081	0.079	0.121	0.142	0.197	0.252	0.253
Paper & publishing	0.077	0.067	0.067	0.049	0.045	0.035	0.026
Professionals	0.070	0.047	0.057	0.050	0.043	0.030	0.037
Sea & canal transport	0.056	0.057	0.058	0.086	0.066	0.080	0.077
Instruments & jewelry	0.053	0.054	0.051	0.050	0.037	0.024	0.023
Chemicals & drugs	0.052	0.019	0.043	0.043	0.006	0.041	0.027
Road transport	0.038	0.031	0.029	0.020	0.021	0.010	0.010
Oil, soap, etc.	0.028	0.017	0.008	0.007	0.008	0.024	0.041
Leather, hair, etc. goods	0.026	0.020	0.025	0.026	0.023	0.023	0.039
Vehicles	0.021	0.012	0.006	0.005	0.033	0.049	0.053
General services	0.020	0.020	0.016	0.022	0.020	0.025	0.026
Tobacco	0.020	0.022	0.010	-0.008	-0.008	0.011	0.016
Earthenware & bricks	0.018	0.026	0.032	0.024	0.015	0.009	0.007
Wood & furniture	0.016	0.015	0.022	0.019	0.015	0.009	0.010
Drinks	0.012	0.011	0.010	0.017	0.008	0.001	-0.002
Food processing	0.009	0.004	0.003	0.001	0.001	0.001	0.001
Clothing, shoes, etc.	0.006	0.007	0.006	0.005	0.008	0.007	0.008
Shopkeepers, salesmen, etc.	0.006	0.001	-0.003	-0.003	-0.004	-0.004	-0.004
Construction	0.004	0.004	0.002	0.003	0.002	0.001	0.002
Merchants, agents, etc.	-0.044	-0.052	-0.047	-0.054	-0.053	-0.055	-0.060
Median	0.026	0.020	0.025	0.022	0.020	0.024	0.026
Mean	0.042	0.039	0.041	0.043	0.042	0.044	0.045

This table reports industry agglomeration in each year based on the index from Ellison & Glaeser (1997). This approach adjusts for the size of plants in an industry using an industry Herfindahl index. We construct these Herfindahl indices using the firm size data reported in the 1851 Census and apply the same Herfindahl for all years, since firm-size data are not reported in later Censuses. This may introduce bias for some industries, such as shipbuilding, where evidence suggests that the average size of firms increased substantially over the study period. Some analysis industries are not included in this table due to lack of firm size data.

Table 3.4: Industry agglomeration patterns excluding London

Industry	1851	1861	1871	1881	1891	1901	1911
Shipbuilding	0.154	0.166	0.161	0.158	0.162	0.184	0.162
Instruments & jewelry	0.152	0.172	0.145	0.163	0.093	0.056	0.045
Sea & canal transport	0.150	0.175	0.161	0.199	0.173	0.164	0.143
Oil, soap, etc.	0.075	0.051	0.031	0.027	0.032	0.066	0.068
Chemicals & Drugs	0.071	0.040	0.044	0.030	0.015	0.025	0.028
Metal & machines	0.056	0.041	0.034	0.034	0.031	0.027	0.026
Textiles	0.046	0.047	0.045	0.050	0.049	0.049	0.049
Mining	0.044	0.052	0.035	0.041	0.043	0.037	0.038
Leather, hair, etc.	0.018	0.019	0.029	0.026	0.026	0.022	0.020
Earthenware & bricks	0.015	0.010	0.008	0.007	0.009	0.010	0.010
Vehicles	0.015	0.023	0.014	0.017	0.046	0.052	0.074
Road transport	0.010	0.013	0.008	0.007	0.005	0.002	0.002
Clothing, shoes, etc.	0.009	0.008	0.010	0.014	0.023	0.017	0.017
Drinks	0.009	0.008	0.009	0.007	0.005	0.003	0.001
Tobacco	0.008	0.003	-0.002	0.018	0.018	0.042	0.072
Shopkeepers, salesmen, et	0.006	0.002	0.001	0.004	0.000	-0.001	-0.003
Food processing	0.006	0.004	0.002	0.002	0.002	0.002	0.003
Wood & furniture	0.006	0.005	0.005	0.003	0.004	0.004	0.004
Paper & publishing	0.004	0.004	0.003	0.004	0.004	0.004	0.005
Construction	0.003	0.002	0.001	0.001	0.001	0.001	0.001
General services	0.003	0.000	-0.001	-0.002	-0.002	-0.002	-0.002
Professionals	-0.006	-0.007	-0.007	-0.007	-0.008	-0.008	-0.008
Merchants, agents, etc.	-0.048	-0.050	-0.052	-0.063	-0.066	-0.067	-0.067
Median	0.010	0.010	0.009	0.014	0.015	0.017	0.017
Mean	0.035	0.034	0.030	0.032	0.029	0.030	0.030

This table reports industry agglomeration in each year based on the index from [Ellison & Glaeser \(1997\)](#). This approach adjusts for the size of plants in an industry using an industry Herfindahl index. We construct these Herfindahl indices using the firm size data reported in the 1851 Census and apply the same Herfindahl for all years, since firm-size data are not reported in later Censuses. Some analysis industries are not included in this table due to lack of firm size data.

Table 3.5: Summary statistics for the cross-industry spillover terms

Main analysis matrices and industry categories					
	Obs.	Mean	SD	Min	Max
$\Delta \ln(L_{ict+1})$	4,253	0.20	0.38	-5.42	3.57
$\ln(L_{ict+1})$	4,253	6.51	1.94	0.00	13.01
$\sum_{k \neq i} IOin_{ki} \ln(L_{kct})$	4,253	7.94	2.45	2.06	19.60
$\sum_{k \neq i} IOout_{ki} \ln(L_{kct})$	4,253	7.70	5.98	0.00	42.77
$\sum_{k \neq i} EMP_{ki} \ln(L_{kct})$	4,253	99.71	42.21	-91.74	190.24
$\sum_{k \neq i} OCC_{ki} \ln(L_{kct})$	4,253	36.00	25.47	-1.10	110.35
Alternative 1841 IO matrix with aggregated industries					
	Obs.	Mean	SD	Min	Max
$\Delta \ln(L_{ict+1})$	2,222	0.20	0.34	-3.31	3.57
$\ln(L_{ict+1})$	2,222	7.34	1.89	1.79	13.44
$\sum_{k \neq i} IOin1841_{ki} \ln(L_{kct})$	2,222	2.72	2.79	0.00	12.10
$\sum_{k \neq i} IOout1841_{ki} \ln(L_{kct})$	2,222	3.94	3.88	0.00	11.76
$\sum_{k \neq i} EMP_{ki} \ln(L_{kct})$	2,222	49.39	24.75	-29.16	95.01
$\sum_{k \neq i} OCC_{ki} \ln(L_{kct})$	2,222	24.83	16.46	-0.66	66.92

Note: We report cross-city summary statistics for 1861-1911 because we only report instrumented cross-city regression results in the main text, which means that 1851 is used only to construct lagged values. For the others, we report summary statistics using the full 1851-1911 period since we report both OLS and instrumented results.

It is also useful to look at the correlation between the cross-industry terms included in Eq. 2.18. These correlations are described in Table 3.6 below for the set of cross-industry terms used in the main analysis. In general we can see that the correlations between these variables are not too high, with the greatest correlation showing up between the IOin and IOout terms.

Table 3.6: Correlations between cross-industry terms used in the main analysis

	IOin	IOout	EMP	OCC
IOin	1.0000			
IOout	0.5478	1.0000		
EMP	0.1218	0.1789	1.0000	
OCC	0.0741	-0.1783	0.3035	1.0000

Empirical approach appendix

Monte Carlo simulations

We use Monte Carlo simulations to assess how well our estimation strategy performs in datasets displaying the size and characteristics of our data. The basic idea is to generate datasets that mimic our real data, but obtained from a data generating process (DGP) with known parameter values. We then apply our estimation strategy to these placebo data sets, recover parameter estimates, and compare them to the estimates obtained in the true data. This allows us to assess the ability of our estimation strategy to obtain unbiased results and accurate confidence intervals.

We begin by estimating our baseline regression specification, Eq. 2.18, in order to obtain a set of industry-year effects ($\hat{\phi}_{it}$), city-year ($\hat{\theta}_{ct}$) effects, and estimated residuals $\hat{\varepsilon}_{cit}$. These ingredients will be used to simulate new datasets in which the city-year and industry-year effects are held constant at the estimated values, and the error terms are drawn from a multivariate Normal distribution whose parameters are computed using the estimated residuals.

Step 1 – constructing the simulated error term

We want to generate a simulated error vector that displays correlation within the city-year (CY), industry-year (IY) and city-industry (CI) dimensions but is uncorrelated across these dimensions. In other words, we need to draw entire vectors of errors ε_{cit} from a multivariate distribution whose covariance matrix Ω has zeros if two observations do not share any cluster, and non-zeros if they share at least a cluster. We follow [Cameron](#)

et al. (2011) and construct such multi-clustered covariance matrix Ω as the sum of four single-clustered covariance matrices.¹⁷

$$\Omega = \Omega^{CY} + \Omega^{IY} + \Omega^{CI} - 2\Omega^{CIY}$$

Notice that if we sort the observations by a given dimension of clustering x , Ω^x has a block diagonal structure. For example, Ω^{CY} consists of blocks of zeros if the corresponding observations are not in the same city-year cluster, and blocks along the diagonal with elements potentially different from zero if the corresponding observations are from the same city-year pair. We denoted these non-zero submatrices by W^{CY} and assume that they are identical across clusters. Therefore the typical element of W^{CY} is $\sigma^{ij} = cov(\varepsilon_{cit}, \varepsilon_{cjt}) \neq 0$.

We use the estimated residuals $\hat{\varepsilon}_{cit}$ from the baseline specification to construct the elements of each submatrix W^x . For instance, taking any two industries i and j , we set $\hat{\sigma}^{ij} = \frac{1}{\#CY} \sum_{CY} \hat{\varepsilon}_{cit} \hat{\varepsilon}_{cjt}$, where $\#CY$ is the number of different city-year pairs. We compute the elements of Ω^{IY} and Ω^{CI} in the same way. We take a different approach to compute the elements of Ω^{CIY} since each cluster has only one observation, i.e. there's a single observation for each triplet city-industry-year. All the diagonal elements of Ω^{CIY} are set to the mean squared residual, i.e. $\hat{\sigma}^{cit} = \hat{\sigma} = \frac{1}{N} \sum_{CIY} \hat{\varepsilon}_{cit}^2$, where N is the number of observations. The off-diagonal elements of Ω^{CIY} are zeros.¹⁸

We draw 500 vectors of error terms from the multivariate distribution $N(0, \Omega)$ and rescale each vector so that it has exactly the same mean (zero) and variance as the orig-

¹⁷Following [Cameron et al. \(2011\)](#)'s notation, with three non-nested dimensions of clustering (denoted by A, B, C) the correct formula to compute a multi-clustered covariance matrix is $\Omega^{ABC} = \Omega^A + \Omega^B + \Omega^C - \Omega^{A \cap B} - \Omega^{A \cap C} - \Omega^{B \cap C} + \Omega^{A \cap B \cap C}$ where, for instance, the entries of Ω^A are non-zero if two observations share the same cluster along a single dimension A , while the entries of $\Omega^{A \cap B}$ are non-zero if two observations share the same cluster defined by the intersection of A and B . In our application, notice that $\Omega^{CY \cap IY} = \Omega^{CY \cap CI} = \Omega^{IY \cap CI} = \Omega^{CIY}$, therefore the formula above collapses to four distinct terms only.

¹⁸As noted in [Cameron et al. \(2011\)](#), multi-clustered covariance matrices are not guaranteed to be positive semidefinite. When that happens, as in our case, such Ω cannot be used by a random number generator. Our solution is to replace Ω with the nearest positive semidefinite matrix computed using Matlab routine `nearestSPD`.

inal residuals. The result of this procedure is a simulated error term $\tilde{\varepsilon}^{SIM}$ that displays correlated errors along the city-year, industry-year and city-industry dimensions with a variance matching that of the original estimated error term.

Part 2: Simulating the data

The next step in our procedure involves simulating a new set of data with the same dimensions as the original data and with known within-industry and cross-industry spillover parameters.

In order to generate a simulated growth rate for the first period we begin with the level of initial city-industry employment from the data and use Eq. 2.18 to compute a simulated employment growth rate for each city-industry. So, for example, if we let $\beta_1 = 0.05$ and all other β terms and τ_{ii} terms to zero then growth rate of employment in city c and industry i is:

$$\tilde{g}_{ic1} = 0.05 \sum_{k \neq i} IOin_{ki} \ln(L_{kc0}) + \hat{\phi}_{i1} + \hat{\theta}_{c1} + \tilde{\varepsilon}_{ic1}^{CY-IY-CI} \quad (3.7)$$

where $IOin_{ki}$ is the actual input-output weight observed in the data. The shifters $\hat{\phi}_{it}$ and $\hat{\theta}_{ct}$ are kept constant across simulations at the values estimated in the initial regression.

We use this simulated growth rate to obtain L_{kc1} , the level of city-industry employment in the following period, which is then fed back into Eq. 3.7 to obtain L_{kc2} , and so on. We repeat the process until we generate a level of employment for each city-industry-year triplet observed in the data. This procedure delivers a simulated dataset that by construction has the desired clustered error structure and the same number of observations as the original data.

Step 3: Results

We follow this procedure to generate 500 datasets that look like the true data, but that are generated using a data generating process with known τ_{ii} and β parameters. Specifically, for the plots below we set all of the τ_{ii} and β parameters to zero (though

we have also explored alternative non-zero values). We apply our estimation strategy (as in Table 2.2 Column 6) to each of the simulated data sets and obtain a distribution of estimated τ and β parameters.

Figure 3.9 displays the mean, 90% and 95% confidence intervals for the distribution of estimated parameters when the true underlying spillover parameters are set to zero. We can see that our estimators are unbiased. Similar unbiased patterns appear when we use alternative non-zero parameters for either the within or cross-industry spillover terms.

We can compare the distribution of estimated coefficients coming out of this counterfactual DGP with the estimates obtained using the real dataset. This allows us to assess the likelihood of observing the real dataset and the corresponding estimates under the null hypothesis that all parameters are zeros. This method provides us with an alternative way to do hypothesis testing that does not rely on our multi-dimensional clustered standard errors.

Figure 3.10 plots the distribution of estimated $IOin$ parameters obtained using the 500 simulated data sets, as well as the estimate obtained from the true data. These results suggest that obtaining the point estimate for $IOin$ of 0.0587 that we got from the true data (Table 2.2, Column 6) is extremely unlikely when the true parameter value is zero. The implied p-value is 0.00 and the coefficient is significantly different from zero at the 1% level.

Table 3.7 presents the similar results for all the other coefficients of interest and confirms the significance levels of our baseline results from Column 6 of Table 2.2. This is reassuring because one may wonder whether our dataset is sufficiently large to consistently estimate all the parameters of interest, especially given that the observations are potentially correlated across multiple dimensions.

Discussion

These monte carlo results can help us assess how well our approach performs on simulated data sets sharing the same size and variance as the data used in our main analysis.

However, this procedure comes with obvious limitations. In particular, we are assuming that the model is correctly specified and that the error terms are clustered in a particular way. Thus, this simulation cannot be used to assess how well our procedure performs under alternative data generating processes or when standard errors display alternative clustering patterns.

Figure 3.9: Estimates and C.I.s from simulated results when all spillover parameters are zero

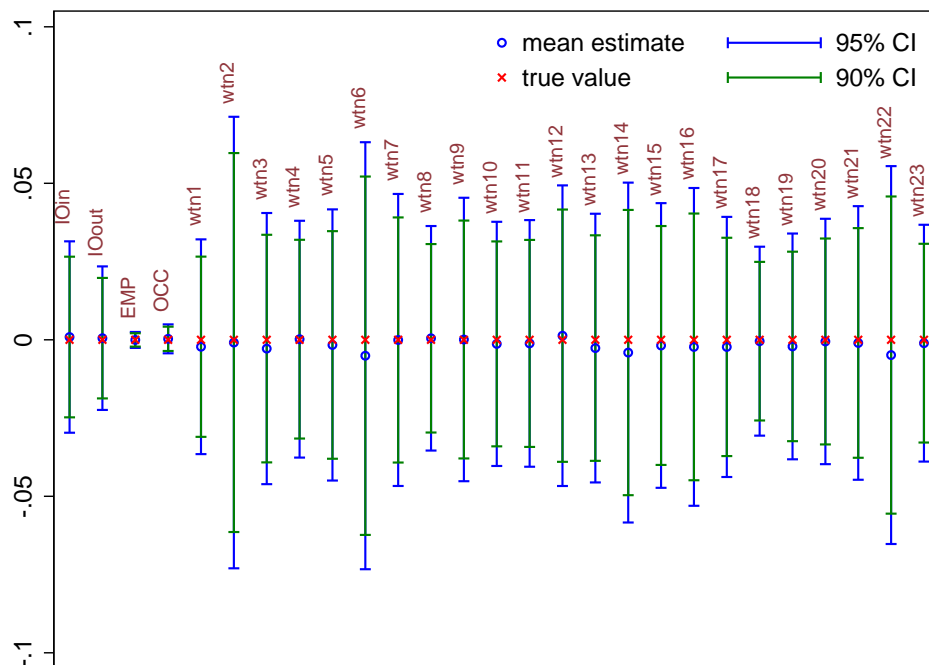


Figure 3.10: Simulated results with all parameters are set to zero vs. *IOin* estimate on real data

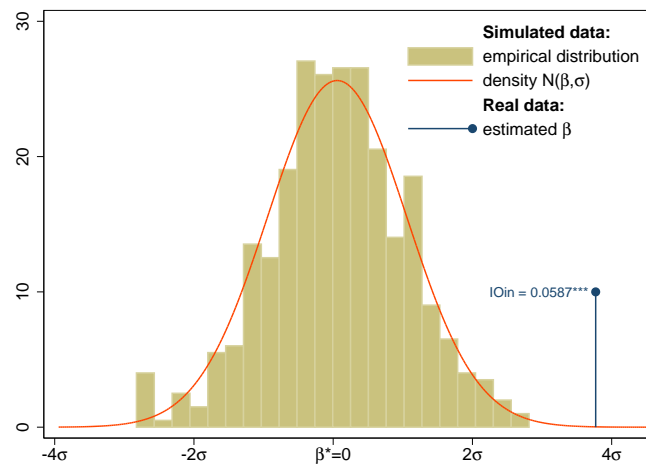


Table 3.7: Simulated results with all parameters are set to zero vs. parameter estimates from true data

Variable	Simulated Data		True Data	
	Mean	Std. Dev.	Coef.	p-value
EMP	0	.001	.002	.213
IOin	.001	.016	.059	0.00
IOout	.001	.012	-.019	.11
OCC	0	.002	.006	.014
wtn1	-.002	.018	-.083	0.00
wtn2	-.001	.037	-.006	.865
wtn3	-.003	.022	-.038	.083
wtn4	0	.019	-.083	0.00
wtn5	-.002	.022	-.035	.117
wtn6	-.005	.035	-.052	.137
wtn7	0	.024	-.038	.114
wtn8	0	.018	.002	.892
wtn9	0	.023	-.05	.031
wtn10	-.001	.02	-.126	0.00
wtn11	-.001	.02	.04	.049
wtn12	.001	.025	-.034	.163
wtn13	-.003	.022	-.083	0.00
wtn14	-.004	.028	-.104	0.00
wtn15	-.002	.023	-.043	.066
wtn16	-.002	.026	-.075	.004
wtn17	-.002	.021	-.069	.001
wtn18	0	.015	-.035	.024
wtn19	-.002	.018	-.034	.067
wtn20	-.001	.02	-.036	.069
wtn21	-.001	.022	-.064	.004
wtn22	-.005	.031	-.145	0.00
wtn23	-.001	.019	-.053	.006

For each of the key explanatory variables, the first two columns of this table present the mean and standard deviation of the distribution of coefficient estimates obtained from applying our estimation strategy to 500 simulated datasets where the data have been generated with all spillover parameter values set to zero. Column 3 presents the coefficients estimated using the true data (as in Table 2.2, Column 6). Column 4 presents the p-value implied by comparing the coefficients estimated using the true data to the distribution of coefficient estimates obtained from the simulated data.

KP test appendix

The standard errors in all of our main regressions are clustered along multiple dimensions. When using 2sls regressions, it is useful to be able to calculate the [Kleibergen](#)

& Paap (2006) test statistics for under- and weak-identification using the appropriately clustered covariance matrix. The KP statistics can easily be computed using existing Stata routines, but only for up to two non-nested dimensions of clustering (Kleibergen (2010)). None of these routines can handle a higher number of clusters so we developed our own package, which we will make available to the benefit of other researchers.

Our strategy builds on Thompson (2011) and Cameron *et al.* (2011) to compute a multi-clustered covariance of the orthogonality condition for any number of clusters. We then use a modified version of the Stata program *ranktest* to compute the appropriate KP statistics based on this covariance matrix. It can be verified that our program exactly reproduces the rk statistic (under-identification) and Wald statistic computed by *ranktest* in the case of two clusters. The weak-identification test statistic is then computed by transforming the Wald statistic into an F statistic. Notice that the value of our F statistic does not exactly match the one computed by *ivreg2* due to the very small differences in the small sample adjustment.

Results appendix

Robustness of results to dropping cities or industries

Figure 3.11 presents histograms of t-statistics for each cross-industry term obtained from running regressions equivalent to Column 6 of Table 2.2, where in each regression a different city is dropped from the dataset. This allows us to assess the extent to which our results are robust to changes in the set of cities included in the analysis. These results indicate that our estimates are not sensitive to dropping individual cities from the analysis database.

Figure 3.11: Robustness to dropping one city at a time – distribution of t-statistics

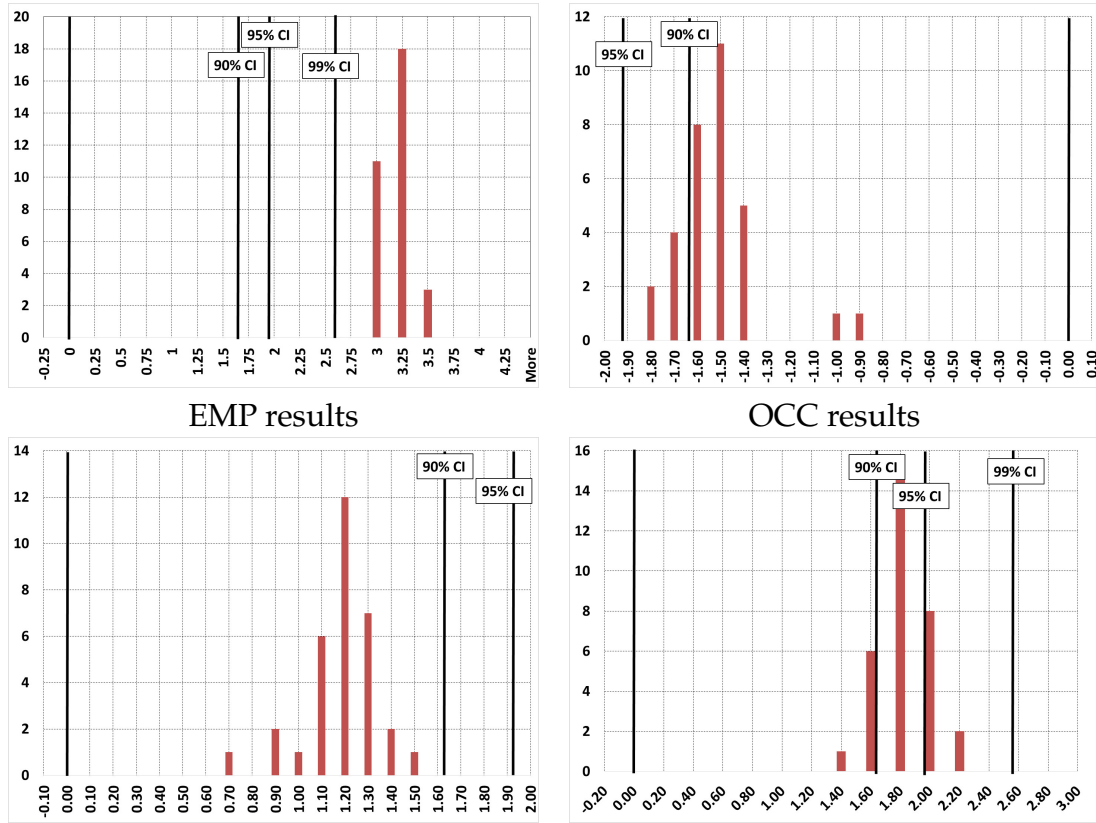
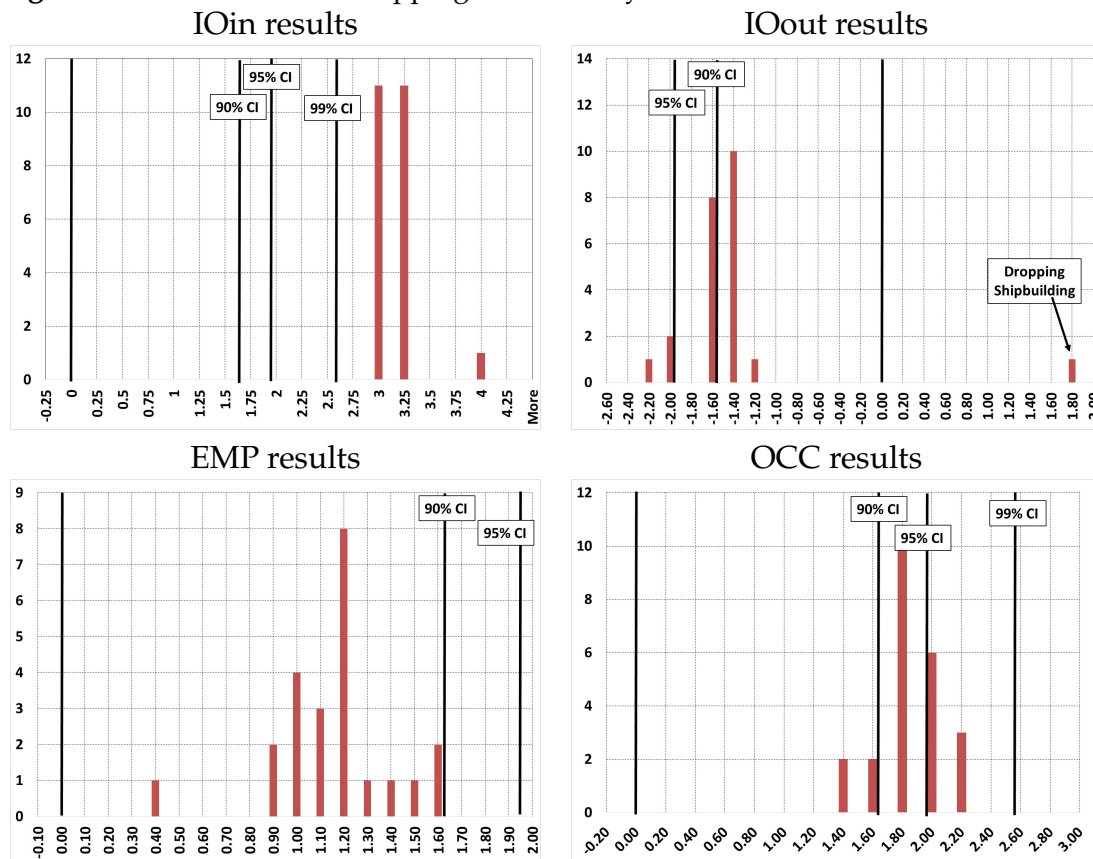


Figure 3.12 presents histograms of t-statistics for each cross-industry term obtained from running regressions equivalent to Column 6 of Table 2.2, where in each regression a different industry is dropped from the dataset. This allows us to assess the extent to which our results are robust to changes in the set of industries included in the analysis. We can see that in general our estimated coefficients are not sensitive to dropping individual industries. However, this does not apply when looking at the IO out coefficient. The top-right graph shows that when we drop shipbuilding from the data, the IO out coefficient changes substantially. In particular, the estimated coefficient changes from negative and occasionally statistically significant to positive and not statistically significant. This suggests that the negative coefficient estimated on the IO out coefficient is driven entirely by the Shipbuilding industry. This is an unusual industry because presumably it can only operate in coastal cities or those with access to a major navigable river. Thus, the IO

out results obtained when dropping this industry seem more reasonable. These results suggest that in general the impact of local customers is weakly positive.

Overall, the results in Figure 3.12 indicate that our estimates are much more sensitive to dropping industries than they are to dropping cities. This suggests that heterogeneity across industries is more important than heterogeneity across cities.

Figure 3.12: Robustness to dropping one industry at a time – distribution of t-statistics



Robustness: weighted by initial city-industry employment

Table 3.8 presents additional results in which each observation has been weighted by initial city-industry employment. Weights for each observation are based on employment in the city-industry at the beginning of each period.¹⁹ The estimated results on the IOin

¹⁹This weighting approach is slightly different than the approach used in some previous drafts of this paper. In previous drafts we often weighted all observations by city-industry employment in 1851. In this draft we allow the weights to adjust over time as cities and industries grow. We believe that this is a better approach because it does not over-weight the industries or cities which were large in 1851 but were much

term in these weighted regressions is very similar to that found in the unweighted regressions presented in the main text. A difference between the weighted and unweighted results appears for the IOout term, which appears to be more negative in the weighted results. However, this coefficient continues to be highly sensitive to the set of industries included in the analysis, so we do not interpret this as a strong result. Another difference is that the agglomeration benefits between industries using similar occupation disappears when looking at weighted regressions. This suggests that labor market pooling benefits may be larger when the local size of an industry is small.

Table 3.8: Weighted regression results with all cross-industry spillover channels

	(1)	(2)	(3)	(4)	(5)	(6)
IOin1	0.0210** (0.0083)	0.0260*** (0.0089)	0.0276*** (0.0097)	0.0510*** (0.0141)	0.0484*** (0.0136)	0.0551*** (0.0153)
IOout1	-0.0130*** (0.0043)	-0.0153*** (0.0045)	-0.0160*** (0.0046)	-0.0324*** (0.0088)	-0.0356*** (0.0081)	-0.0391*** (0.0079)
EMP	-0.0001 (0.0003)	-0.0001 (0.0003)	-0.0001 (0.0003)	0.0002 (0.0009)	0.0006 (0.0009)	0.0005 (0.0009)
OCC	-0.0003 (0.0007)	0.0001 (0.0007)	0.0001 (0.0008)	-0.0019 (0.0018)	-0.0011 (0.0020)	-0.0013 (0.0020)
Within	-0.0156* (0.0088)	-0.0142 (0.0099)	-0.0147 (0.0098)			
Observations	4,253	3,544	3,539	4,253	3,539	3,539
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
wtn	homog	homog	homog	heter	heter	heter
KP under		24.86	25.45		22.09	24.52
KP weak id.		4677.9	858.61		52.36	35.68

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Heterogeneous regressors *within* are included in Columns 4-6 but not displayed. City-year and industry-year effects are included in all regressions but not displayed. 2SLS regressions use lagged instruments. Note that the number of observations falls for the instrumented regressions in columns 3-6 because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*. “KP under id.” denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). “KP weak id.” denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic. Weights for each city-industry observation are based on employment in the city-industry at the beginning of each period.

less important 60 years later.

Heterogeneous effects

In this section we look at heterogeneity in the pattern of cross-industry and within-industry effects across different industries. We begin by considering heterogeneous cross-industry effects. Specifically, we run two alternative versions of Equation 2.18,

$$\Delta \ln(L_{ict+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \beta^i \sum_{k \neq i} \text{CONNECT}_{ki} \ln(L_{kct}) + \theta_{ct} + \chi_{it} + e_{ict} \quad (3.8)$$

$$\Delta \ln(L_{i \neq kct+1}) = \tilde{\tau}_{ii} \ln(L_{ict}) + \beta^k \text{CONNECT}_{ki} \ln(L_{kct}) + \theta_{ct} + \chi_{it} + e_{ict} \quad (3.9)$$

where CONNECT_{ki} is one of our four measures of cross-industry connections. Equation 3.8 allows us to estimate industry-specific coefficients β^i describing how much each industry i benefits from cross-industry connections. This specification can be estimated using the same approach as was used for our baseline regressions. Using Equation 3.9, we estimate industry-specific coefficients β^k that reflect the extent to which industry k generates spillovers that benefit other industries. Estimating this value requires a different approach to avoid conflating the within and between impact of industry k when estimating β^k . Specifically, we run separate regressions corresponding to Equation 3.9 for each industry k . In each of these regressions, only employment in industry k (interacted with a cross-industry connection measure) is included as an explanatory variable and observations from industry k are not included in the dependent variable.

Once the industry-specific β^i and β^k terms are estimated, we compare them to available measures of industry characteristics: firm size in each industry, the share of output exported, the share of output sold to households, the industry labor cost share, and the industry intermediate cost share. In each case we run a simple univariate regression where the dependent variable is the estimated industry-specific cross-industry spillover coef-

ficient and the independent variable is one of the industry characteristics. Univariate regressions are used because we are working with a relatively small number of observations. These results can provide suggestive evidence about the characteristics of industries that produce or benefit from different types of cross-industry spillovers, but because of the small sample size we will not be able to draw any strong conclusions.

Table 3.9 describes the characteristics of industries that *benefit from* cross-industry connections. In rows 1-2, we see evidence that small firm size in an industry is associated with more cross-industry spillover benefits, but this pattern is not statistically significant at standard confidence levels. The only strong result coming out of this table is that industries that benefit from connections to other local industries with similar labor pools tend to have a larger labor cost share relative to overall industry sales, as well as a smaller intermediate cost share. This seems like a very reasonable result which provides some additional confidence that the estimates we have obtained are reasonable.

Table 3.9: Features of industries that benefit from each type of cross-industry spillover

Spillovers channel:	Coefficients from univariate regressions			
	DV: Estimated industry-specific β^i coefficient			
	IO-in	IO-out	EMP	OCC
Average firm size	-0.273 (0.330)	-1.100* (0.608)	-0.0303 (0.0328)	-0.235 (0.945)
Median worker's firm size	-0.0217 (0.0390)	-0.114 (0.0726)	-0.00234 (0.00389)	-0.0333 (0.111)
Share of industry output exported abroad	-0.0741 (0.104)	-0.0695 (0.198)	-0.0150 (0.0107)	-0.157 (0.284)
Share of industry output sold to households	0.0612 (0.0459)	0.161 (0.0932)	0.00689 (0.00485)	0.0975 (0.128)
Labor cost/output ratio	-0.101 (0.146)	-0.337 (0.274)	-0.0084 (0.00993)	0.413** (0.186)
Intermediate cost/output ratio	0.0092 (0.107)	0.143 (0.195)	-0.00059 (0.00733)	-0.364*** (0.122)

Estimated coefficients from univariate regressions. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The dependent variable in each regression is the estimated β^i coefficient from Eq. 3.8. Firm size data comes from the 1851 Census of Population. The share of industry output exported or sold to households is from the 1907 input-output table. The labor cost share is constructed from industry wage bills from the 1907 Census of Manufactures. The intermediate cost share is based on the 1907 input-output table. We do not report robust standard errors because these generate smaller confidence intervals, probably due to small-sample bias. We have also explored regressions in which we weight results by the inverse of the standard error of each estimated within-industry coefficient in order to account for the precision of those estimates and these deliver similar results.

Table 3.10 describes the characteristics of industries that *produce* cross-industry connections. These results also suggest that industries with smaller firm sizes produce more beneficial cross-industry spillovers, but again, these results are not statistically significant. As before, we observe is that industries with smaller intermediate cost share relative to overall sales produce fewer cross-industry benefits to occupationally similar industries. There is some evidence that this may be linked to the importance of labor in firm inputs.

Table 3.10: Features of industries that produce each type of cross-industry spillover

Spillovers channel:	Coefficients from univariate regressions			
	DV: Estimated industry-specific β^k coefficient			
	IO-in	IO-out	EMP	OCC
Average firm size	-1.250 (1.060)	-3.125 (6.269)	0.00417 (0.180)	-1.809 (2.048)
Median worker's firm size	-0.140 (0.125)	-0.543 (0.731)	-0.00288 (0.0211)	-0.189 (0.242)
Share of industry output exported abroad	-0.0495 (0.349)	-0.808 (1.934)	-0.0121 (0.0556)	-0.550 (0.623)
Share of industry output sold to households	0.0013 (0.175)	0.0045 (0.879)	-0.0119 (0.0250)	0.483* (0.266)
Labor cost/output ratio	0.0154 (0.547)	1.224 (3.101)	-0.0296 (0.0505)	0.494 (0.341)
Intermediate cost/output ratio	-0.305 (0.354)	-0.314 (2.191)	0.0176 (0.0356)	-0.493** (0.219)

Estimated coefficients from univariate regressions. The dependent variable in each regression is the estimated β^k coefficient from Eq. 3.9. Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Firm size data comes from the 1851 Census of Population. The share of industry output exported or sold to households is from the 1907 input-output table. The labor cost share is constructed from industry wage bills from the 1907 Census of Manufactures. The intermediate cost share is based on the 1907 input-output table. We do not report robust standard errors because these generate smaller confidence intervals, probably due to small-sample bias. We have also explored regressions in which we weight results by the inverse of the standard error of each estimated within-industry coefficient in order to account for the precision of those estimates and these deliver similar results.

Next, we undertake a similar exercise with our estimated within-industry coefficients. In Table 3.11 we consider some of the industry characteristics that may be related to the range of different within-industry spillover estimates we observe. Columns 1-2 focus on the role of firm size using two different measures. We observe a positive relationship between firm size in an industry and the strength of within-industry spillovers, but this results is not statistically significant due to the small number of available observations. There is also weak evidence that more labor intensive industries benefit more from within-industry spillovers.

Table 3.11: Features of industries that benefit from within-industry spillovers

DV: Estimated industry-specific within-industry spillover coefficients						
Average firm size	0.172					
	(0.198)					
Median worker's firm size		0.0176				
		(0.0233)				
Exports share of industry output			-0.0313			
			(0.0705)			
Households share of industry output				-0.0247		
				(0.0307)		
Labor cost/output ratio					0.0699	
					(0.0875)	
Intermediate cost/output ratio						-0.0304
						(0.0663)
Observations	20	20	20	20	15	15
R-squared	0.040	0.031	0.011	0.035	0.047	0.016

Standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. The number of observations varies because the explanatory variables are drawn from different sources and are not available for all industries. The within coefficients come from the specification used in Column 6 of Table 2.2. Firm size data comes from the 1851 Census of Population. The export's and household's share of industry output come from the input-output table. Total labor cost and total output values come from the 1907 Census of Production. Intermediate cost is constructed based on data from the 1907 Input-Output matrix. We do not report robust standard errors because these generate smaller confidence intervals, probably due to small-sample bias. We have also explored regressions in which we weight results by the inverse of the standard error of each estimated within-industry coefficient in order to account for the precision of those estimates and these deliver similar results.

Robustness: Alternative functional forms

In this table we replace the logarithms on the right-hand side of the estimating equation with plausible alternative functional forms based on either the second root or fifth root. These results show that adjusting the functional form in this way has little impact on the estimated results.

Table 3.12: Regression results with alternative functional forms

FF:	Square root			Fifth root		
	(1)	(2)	(3)	(4)	(5)	(6)
IOin	0.0013** (0.0006)	0.0012** (0.0005)	0.0013** (0.0005)	0.0598*** (0.0209)	0.0526** (0.0205)	0.0561*** (0.0211)
IOout	-0.0003 (0.0004)	-0.0004 (0.0004)	-0.0005 (0.0004)	-0.0133 (0.0142)	-0.0157 (0.0135)	-0.0190 (0.0137)
EMP	0.0000 (0.0001)	0.0001 (0.0001)	0.0001 (0.0001)	0.0015 (0.0022)	0.0034* (0.0018)	0.0024 (0.0019)
OCC	0.0003*** (0.0001)	0.0003* (0.0001)	0.0003** (0.0001)	0.0085** (0.0037)	0.0078* (0.0041)	0.0076* (0.0041)
Observations	4,253	3,544	3,544	4,253	3,544	3,544
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
KP under		13.45	14.47		19.69	22.06
KP weak		198.94	68.59		55.61	34.2

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. A full set of industry-specific *within* terms, industry-year and city-year effects are included in all regressions but not displayed. Regressions in Columns 2 and 4 instrument the *within* terms with lagged values. Regressions in Columns 3 and 5 instrument both the *within* and *between* terms with lagged values. Acronyms: wtn = *within*, btn = *between*. “KP under” denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). “KP weak” denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic.

Robustness: Alternative connections matrices

Next, we revisit the analysis using some alternative measures of inter-industry connections. In particular, we use an alternative matrix of input-output connections constructed by Horrell *et al.* (1994) for Britain in 1841. Generating results with this alternative matrix, which comes from before the study period, can help address concerns that the results we find are dependent on the specific set of matrices we consider or are due to a process of endogenous inter-industry connection formation. The cost of using this matrix is that we are forced to work with a smaller set of 12 more aggregated industry categories.²⁰

Because we are now working with a smaller number of industry categories, we focus on regressions that incorporate one spillover channel at a time. Table 3.13 describes the results. As in the main results, we observe positive effects occurring through the IOin

²⁰The industry categories are: “Mining & quarrying,” “Food, drink & tobacco,” “Metals & Machinery,” “Oils, chemicals & drugs,” “Textiles, clothing & leather goods,” “Earthenware & bricks,” “Other manufactured goods,” “Construction,” “Gas & water,” “Transportation,” “Distribution,” and “All other services.”

channel and these results are generally statistically significant. There is also evidence that industries benefited from the presence of local buyers, but this result is clearly sensitive to the underlying set of industries used, so it should be interpreted with some caution. There is also some evidence of benefits through the presence of occupationally similar local industries.

Table 3.13: Alternative matrix regressions with one channel at a time

	(1)	(2)	(3)	(4)	(5)	(6)
IOin1841	0.0241* (0.0126)	0.0194 (0.0134)	0.0259* (0.0149)			
IOout1841				0.0593** (0.0299)	0.0580** (0.0254)	0.0651*** (0.0248)
Observations	2,222	1,850	1,850	2,222	1,850	1,850
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
KP under id.		14.92	17.16		12.81	14.88
KP weak id.		49.14	22.63		35.55	33.17
	(7)	(8)	(9)	(10)	(11)	(12)
EMP	0.0049 (0.0033)	0.0064* (0.0037)	0.0044 (0.0027)			
OCC				0.0020 (0.0074)	0.0077 (0.0061)	0.0063 (0.0061)
Observations	2,222	1,850	1,850	2,222	1,850	1,850
Estimation	ols	2sls	2sls	ols	2sls	2sls
Instrumented	none	wtn	wtn-btn	none	wtn	wtn-btn
KP under id.		12.86	13.96		6.43	16.71
KP weak id.		42.89	39.48		19.94	61.55

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. A full set of *within* regressors, city-time and industry-time effects are included in all regressions but not displayed. 2SLS regressions use lagged instruments. Note that the number of observations falls for the instrumented regressions because the instruments require a lagged employment term. Thus, data from 1851 are not available for these regressions. Acronyms: wtn = *within*, btn = *between*. "KP under id." denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). "KP weak id." denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic.

Robustness: Cross-city effects

There is substantial variation in the proximity of cities in our database to other nearby cities. Some cities, particularly those in Lancashire, West Yorkshire, and the North Midlands, are located in close proximity to a number of other nearby cities. Others, such as Norwich, Hull, and Portsmouth are relatively more isolated. In this section, we extend our analysis to consider the possibility that city-industry growth may also be affected by forces due to other nearby cities.

We consider two potential channels for cross-city effects. First, industries may benefit from proximity to consumers in nearby cities. This *market potential* effect has been suggested by [Hanson \(2005\)](#), who finds that regional demand linkages play an important role in generating spatial agglomeration using modern U.S. data. Second, industries may benefit from spillovers from other industries in nearby towns, through any of the channels that we have identified. We analyze these effects using the more detailed industry categories from Section 2.5.

We begin our analysis by collecting data on the distance (as the crow flies) between each of the cities in our database, which we call $distance_{ij}$. Using these, we construct a measure for the remoteness of one city from another $d_{ij} = \exp(-distance_{ij})$.²¹ Our measures of market potential for each city is then,

$$MP_{ct} = \ln \left(\sum_{j \neq c} POP_{jt} * d_{cj} \right).$$

where POP_{jt} is the population of city j . This differs slightly from Hanson's approach, which uses income in a city instead of population, due to the fact that income at the city level is not available for the period we study.

We also want to measure the potential for cross-industry spillovers occurring across cities. We measure proximity to an industry i in other cities as the distance-weighted sum

²¹This distance weighting measure is motivated by [Hanson \(2005\)](#). We have also explored using $d_{ij} = 1/distance_{ij}$ as the distance weighting measure and this delivers similar results.

of log employment in that industry across all other cities. Our full regression specification, including both cross-city market potential and spillover effects, is then,

$$\begin{aligned}
\Delta \ln(L_{ict+1}) = & \tilde{\tau}_{ii} \ln(L_{ict}) \\
& + \beta_1 \sum_{k \neq i} IOin_{ki} \ln(L_{kct}) + \beta_2 \sum_{k \neq i} IOout_{ki} \ln(L_{kct}) \\
& + \beta_3 \sum_{k \neq i} EMP_{ki} \ln(L_{kct}) + \beta_4 \sum_{k \neq i} OCC_{ki} \ln(L_{kct}) \\
& + \beta_5 \left[\sum_{k \neq i} IOin_{ki} \sum_{j \neq c} d_{jc} * \ln(L_{kjt}) \right] + \beta_6 \left[\sum_{k \neq i} IOout_{ki} \sum_{j \neq c} d_{jc} * \ln(L_{kjt}) \right] \\
& + \beta_7 \left[\sum_{k \neq i} EMP_{ki} \sum_{j \neq c} d_{jc} * \ln(L_{kjt}) \right] + \beta_8 \left[\sum_{k \neq i} OCC_{ki} \sum_{j \neq c} d_{jc} * \ln(L_{kjt}) \right] \\
& + \beta_9 MP_{ct} + \log(WORKpop_{ct}) + \theta_c + \chi_{it} + \epsilon_{ict}.
\end{aligned}$$

One difference between this and our baseline specification is that we now include city fixed effects (θ_c) in place of city-year effects because city-year effects would be perfectly correlated with the market potential measure. To help deal with city-size effects, we also include the log of $WORKpop_{ct}$, the working population of city c in period t . To simplify the exposition and in analogy with the previous section, we will refer to the cross-city term $\sum_{k \neq i} IOin_{ki} \sum_{j \neq c} d_{jc} * \ln(L_{kjt})$ as $IOin * d$, and similarly for the other cross-city terms $IOout * d$, $EMP * d$, and $OCC * d$.

The results generated using this specification are shown in Table 3.14. The first thing to take away from this table is that our baseline results are essentially unchanged when we include the additional cross-city terms. The city employment term in the fifth column reflects the negative growth impact of city size. The coefficients on the market potential measure is always positive but not statistically significant.

Table 3.14: Regression results with cross-city variables

	(1)	(2)	(3)
IOin1	0.0502*** (0.0174)	0.0516*** (0.0185)	0.0500*** (0.0192)
IOout1	-0.0241** (0.0119)	-0.0250** (0.0119)	-0.0255** (0.0119)
EMP	-0.0025 (0.0019)	-0.0028 (0.0019)	-0.0029 (0.0019)
OCC2	0.0057* (0.0034)	0.0049 (0.0034)	0.0048 (0.0034)
City Employment	-0.3263*** (0.0801)	-0.3049*** (0.0776)	-0.3095*** (0.0799)
Market Potential	0.1671 (0.1692)		0.1265 (0.2448)
IOin*dist		-0.0002 (0.0022)	-0.0011 (0.0026)
IOout*dist		-0.0011 (0.0011)	-0.0010 (0.0011)
EMP*dist		0.0002 (0.0001)	0.0001 (0.0001)
OCC*dist		0.0001 (0.0002)	0.0001 (0.0002)
Observations	3,549	3,549	3,549
KP under	19.04	20.66	19.07
KP weak	2.02	2.3	2.02

Multi-level clustered standard errors by city-industry, city-year, and industry-year in parenthesis. Significance levels: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. A full set of *within* regressors, city-time and industry-time effects are included in all regressions but not displayed. All regressions instrument the *within* and *between* regressors with lagged instruments. Acronyms: wtn = *within*, btn = *between*. “KP under” denotes the test statistic for the Lagrange Multiplier underidentification test based on Kleibergen & Paap (2006). “KP weak” denotes the test statistic for a weak instruments test based on the Kleibergen-Paap Wald statistic.

The results do not provide statistically significant evidence that cross-city spillovers matter through any of the channels that we measure. However, these results are imprecisely measured. The coefficients estimated on the *IOin* * *dist* term suggest that a one standard deviation increase in the presence of suppliers in other nearby cities could increase city-industry growth by 6.1-18.3%. The coefficients on the *EMP* term are consistent with effects of a similar magnitude. Thus, we should not rule out important cross-city effects based on these results. However, it is clear that omitted cross-city effects are not driving our findings regarding the importance of within-city cross-industry agglomeration forces.